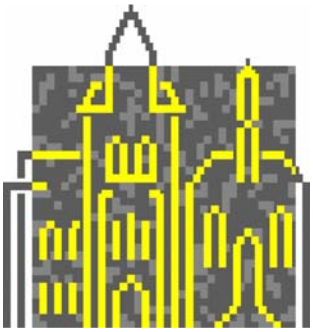

SCHLOSS DAGSTUHL

INTERNATIONAL
CONFERENCE
AND RESEARCH CENTER
FOR COMPUTER SCIENCE

Dagstuhl News

January - December 2004

Volume 7
2006

ISSN 1438-7581

Period: January - December 2004

Frequency: One per year

The International Conference and Research Center for Computer Science is operated by a non-profit organization. Its objective is to promote world-class research in computer science and to host research seminars which enable new ideas to be showcased, problems to be discussed and the course to be set for future development in this field.

Online version: http://www.dagstuhl.de/files/Reports/News/

Associates (2004): Gesellschaft für Informatik e.V. (GI), Bonn
Technical University of Darmstadt
University of Frankfurt
University of Kaiserslautern
University of Karlsruhe
University of Stuttgart
University of Trier
Saarland University

Funded by (2004): The state governments of Saarland
and Rhineland Palatinate

Information: Schloss Dagstuhl Office
Saarland University
P.O. Box 15 11 50
66041 Saarbrücken, Germany
Phone: +49-681-302-4396
Fax: +49-681-302-4397
E-mail: service@dagstuhl.de
http://www.dagstuhl.de/

Welcome

You have in your hands the seventh edition of the "Dagstuhl News", a publication for the members of the Foundation "Informatikzentrum Schloss Dagstuhl", the *Dagstuhl Foundation* for short. As always, we are a bit late, which as always has its reasons in the fact that Dagstuhl keeps us busy and that organizers often are hard of hearing when we ask them to supply us with reports about the results of their Seminars..

You should have received the Dagstuhl calendar for 2007, another gift for you from Dagstuhl in addition to the Dagstuhl News. So, you see, we are trying to improve our relation with our supporters.

We have experienced increasing numbers of guests and overnight stays during the last couple of years. We are approaching the capacity limits, which have been slightly extended by converting some rooms to guest rooms. Also, the number of proposals have increased and are still increasing. Of course, we are happy about these figures as they can be interpreted as proving the unbroken popularity of Dagstuhl. On the other hand, the scheduling problem becomes more and more difficult.

We have now been member of the Leibniz-Gemeinschaft for two years and have received combined federal- and state-funding for one year. The bureaucratic efforts definitely have increased for servicing the needs of such a large community of institutions.

The main part of this volume consists of collected resumees from the Dagstuhl Seminar Reports 2004. We hope that you will find this information valuable for your own work or informative as to what colleagues in other research areas of Computer Science are doing. The full reports for 2004 are on the Web under URL: `http://www.dagstuhl.de/Seminars/04/`

We have switched to publishing online proceedings of our Dagstuhl Seminars on our Dagstuhl Research Online Publication Server (DROPS).
`http://www.dagstuhl.de/publikationen/publikationsserver-drops/`
Authors keep the copyrights to their contributions in order not to harm their rights to submit them to conferences or journals. We hope that the reputation of our Dagstuhl Seminars will make their proceedings a valuable source of information. It encourages us that also external workshops have asked to be hosted on DROPS.

The State and the Activities of the *Dagstuhl Foundation*

The foundation currently has 45 personal members and 7 institutional members.
In 2004, the foundation has supported a few guests with travel grants and a reduction of the Seminar fees. As usual, the supported guests did not have any budget for traveling expenses and could not be financed by Dagtuhl's normal budget. All supported guests were young researchers aged 20-30 years.

Thanks

I would like to thank you for supporting Dagstuhl through your membership in the *Dagstuhl Foundation*. Thanks go to Fritz Müller for editing the resumees collected in this volume.

Reinhard Wilhelm (Scientific Director)

Saarbrücken, January 2007

Contents

Chapter 1

Data Structures, Algorithms, Complexity

1.1 Real Computation and Complexity

Seminar No. **04061** Date **01.02.–06.02.2004**
Organizers: T. Lickteig, K. Meer, L.M. Pardo

The seminar "Real Computation and Complexity" was intended as a meeting place of several tendencies in the complexity analysis of algorithms in real computation. One main idea therefore was to bring together scientists with rather different backgrounds such as numerical analysis, symbolic computing, real and complex algebraic geometry, logic, differential algebra and computational complexity. This broadness guaranteed to get a thorough overview of current results, methods and trends in the area. It allowed as well to discuss main problems related to all aspects of real computation and complexity from different perspectives.

The seminar was attended by 43 participants from 14 different countries (Argentina, Belgium, Brazil, Canada, Denmark, Germany, England, France, Israel, Italy, Russia, Spain, Switzerland, USA). During the five days 34 talks were presented.

The main topics of the seminar as addressed either in talks or in informal discussions were the following:

- complexity upper bounds for linear optimization problems;
- models of computation with real numbers and structural transfer results between them;
- complexity issues and algorithmics in symbolic and numeric multivariate polynomial equation solving and elimination theory;
- quantitative aspects in real equation solving;
- algorithmic aspects and quantitative estimates in differential equation solving;

- fast evaluation of polynomial and analytic functions.

The Dagstuhl-Seminar was devoted to honor renowned scientist, complexity theory pioneer and celebrity Arnold Schönhage on the occasion of his 70th birthday in December 2004. There will be a Festschrift for Arnold Schönhage special volume of Journal of Complexity issue of this Dagstuhl seminar.

1.2 Data Structures

Seminar No. **04091** Date **22.02.–27.02.2004**
Organizers: S. Albers, R. Sedgewick, D. Wagner

The design and analysis of algorithms is a fundamental area in computer science. This also involves the development of suitable methods for structuring the data to be manipulated by these algorithms. Hence, algorithms and data structures form a unit, and the right choice of algorithms and data structures is a crucial step in the solution of many problems. For this reason, the design, analysis and implementation of data structures form a classical field of computer science, both in research and teaching.

The Dagstuhl Seminar on Data Structures in 2004 reported on ongoing research on classical data structuring problems as well as classical application areas such a computational geometry. Furthermore many contributions investigated new algorithmic and data structuring problems arising in large networks or in the maintenance of large data sets. As in previous meeting, there was some shift of interest away from theory, e.g., the classical analysis of asymptotic behavior of algorithms, to more practical issues, such as implementation problems and the usefulness of algorithms in practical applications. This is motivated by the fact that more and more researchers in computer science also want to make their results available in form of programs or software packages.

With 48 participants, the attendence was even higher than in previous meetings on the topic. In addition to scientific talks, there were fruitful and stimulating discussions throughout the meeting. We thank the team of Schloss Dagstuhl for their hospitality and support making this successful workshop possible.

1.3 Complexity of Boolean Functions

Seminar No. **04141** Date **28.03.–02.04.2004**
Organizers: J. Hastad, M. Krause, P. Pudlak, I. Wegener

Background

The determination of the complexity of Boolean functions with respect to various hardware models, like Boolean circuits, branching programs or constant layer feedforward neural

networks, is one of the central and classical topics in the theory of computation. It implies the search for efficient implementations for hardware relevant functions as well as proofs of lower bounds which are established by showing that a given function cannot be computed correctly given certain constraints. In this respect a lot of interesting and surprising results have been obtained which in many cases are based on the development of elegant and very nontrivial mathematical techniques. Methods originally intended to better analyze the complexity of Boolean functions have interesting implications in other areas like hardware verification, computational intelligence and cryptography.

The aim of this seminar was to collect the leading experts of Boolean complexity theory, and to present the latest results in this area. Besides this, a main item in the discussion were new challenges like the analysis of fault tolerant circuits and circuits with feedback or the question if classical complexity theoretic results do hold in the quantum computing setting. A number of successfull applications of Boolean complexity methods in other more applied fields like hardware design and verification, algorithmic learning, quantum computing and cryptanalysis of block- and stream ciphers were presented.

Summary

The talks of the seminar covered a wide spectrum of recently studied questions of Theoretical Computer Science related to the complexity of Boolean functions in nonuniform models. Besides talks on new results for the classical problems in circuit complexity theory (Tesson, Lachish, Bro Miltersen, Tesson, Therien, Kiltz, Jukna) we had contribution on new challenges for Boolean complexity theory like analyzing circuits with feedback (Bruck) or fault tolerant circuits (Newman). Other talks were related to efficient hardware implementations for Boolean functions of practical interest like the integer multiplication (Wlfel), memory functions (Andreev), cryptographic functions (Kiltz) or perfect matching (Thierauf). Also areas of uniform complexity theory like Kolmogorov complexity theory (Lee, Buhman, Koucky) and structural complexity theory (Kabanets, Hesse) were present. One of the deepest results in complexity theory is the PCP-theorem, characterizing the power of probabilistic proof checking. There was a very interesting talk (Reingold) on the search for an easier, more combinatorial and circuit-based proof of the PCP-theorem.

A more applied field which is closely connected to Boolean complexity theory is hardware verification. In this context there were presented interesting contributions on the computational power and algorithmic properties of binary decision diagrams (Wölfel, Sieling, Waack) and on the tractability of SAT problems (van Melkebeek, Iwama, Hofmeister). A classical and in many cases very successful approach to the analysis of restricted hardware models is to describe them in terms of appropriate communication games. We heared interesting talks on characterizing regular languages (Tesson) and contextfree languages and the power of deterministic and randomized pushdown automata (Schnitger) in terms of communication complexity.

A comparatively large number of talks were devoted to a further estimation of the power of quantum models of computation (de Wolf, Klauck, Yao, Jacobi, Spalek). A recent focus in this context is to try to answer the question to which extent basic results of classical complexity theory do hold in a quantum setting.

Last but not least we heared about new results in two traditional application fields for Boolean complexity theory, algorithmic learning theory (Simon, Reischuk), and cryptography, especially cryptanalysis of block- and stream ciphers (Armknecht, Lucks).

1.4 Algorithms and Number Theory

Seminar No. **04211** Date **16.05.–21.05.2004**
Organizers: J. Buhler, J. Cremona, M.E. Pohst

This seminar on number-theoretical algorithms and their applications was the fifth on this topic at Dagstuhl over a period of more than 10 years. This time we attracted a record number of 54 participants from 14 countries.

One of the major goals of these seminars has been to broaden interactions between number theory and other areas. For instance, there has been an effort to bring together people developing the theory of efficient algorithms with people actually writing software. There has also been continuing interest in cryptography, and almost a third of the talks were on algebraic curves, most with an eye to applications in cryptography. Since elliptic curves in cryptography seem to be mainly objects of hardware implementations, nowadays, the focus is on higher genus curves and related more sophisticated mathematical objects.

This time we also had a major new topic: algorithmic K-theory which has been rapidly developing over the last few years. Not surprisingly seven talks were given on this subject, several alone on (algorithmic aspects of) logarithmic class groups.

Most of the other talks focused on more classical topics of algorithmic algebraic number theory, with half a dozen on various aspects of solving Diophantine equations. Among the variety of problems considered we just mention the computation of Picard groups and Drinfeld modules, but also quantum computing of unit groups. Several talks were on problems related to the development of number theoretical software.

The variety of topics was stimulating to the audience. The reaction of the participants was very positive and we believe that we succeeded in having an effective meeting that was able to appeal to a broad audience. We made sure to allow for adequate breaks between sessions, and there were many opportunities for discussions that the participants took advantage of. The pleasant atmosphere of Schloss Dagstuhl once again contributed to a very productive meeting.

1.5 Robust and Approximative Algorithms on Particular Graph Classes

Seminar No. **04221** Date **23.05.–28.05.2004**
Organizers: A. Brandstädt, D. Corneil, K. Jansen, J. Spinrad

The aim of this seminar was to bring together experts working on robust as well as on approximative graph algorithms. Given the fast advances in various areas of graph algorithms on particular graph classes that we have witnessed in the past few years, we believe that it was very important to offer researchers in these areas a forum for the exchange of ideas in the relaxed workshop-like atmosphere that Dagstuhl always offers.

There was a strong interaction and healthy exchange of ideas which resulted in successful applications of robust and approximative graph algorithms; in particular, the seminar had the following aims:

- discussing new graph classes where the recognition complexity of a graph class is harder than solving certain basic graph problems on the class;
- new structural insights by studying the modular and homogeneous decomposition as well as treewidth and clique-width of important graph classes, and the application of these results to the design of robust graph algorithms;
- parameterized complexity and its influence to the design of efficient approximative algorithms;
- solving graph problems approximatively by using (non-)linear programming for graph classes where the problems can be solved in polynomial time but the time bound is poor (e.g. network design problems, Maximum Independent Set for perfect graphs).

The most outstanding result of the discussions during the seminar was the proof of the Seese conjecture by combining results of Courcelle and Sang-il Oum.

1.6 Algorithmic Methods for Railway Optimization

Seminar No. **04261** Date **20.06.–25.06.2004**
Organizers: L. Kroon, D. Wagner, F. Wagner, C. Zaroliagis

The main topic of the seminar is algorithmic methods for analyzing and solving problems arising in railway optimization. Special focus will be on the interplay between railway and other public transport systems. Beside algorithmics and mathematical optimization the relevance of formal models and the influence of application aspects for the problem modelling are considered as well.

Problems arising from the validation and improvement of the railway networks as well as other public transport systems will stand in the center of the discussion. Examples of such problems are: Network planning and Evaluation, Line planning, Timetable construction, Rolling stock circulation, Crew scheduling, Shunting, Delay Management or Timetable information. For analyzing and solving these interesting and challenging problems, typically techniques from network and combinatorial optimization and efficient algorithms must be provided. Transposing these into practical and applicable methods and implementations is a task to be undertaken by computer scientists. However, the high complexity of transport

systems requires a coordinated interdisciplinary effort from mathematicians, computer scientists, traffic engineers, and people from railway companies. Actually, we are expecting intensive knowledge exchange between those groups during the seminar. Aspects to be discussed during the seminar are for example:

- knowledge exchange about problems and methods applied so far in different national railway companies;
- adaptation of techniques and methods already developed according to the requirements and conditions of different national railway systems as well as other transport systems like road traffic or airline transport;
- study of new problems arising from the integration of national railway systems in a European context;
- study of problems arising from the coordination of different transport systems.

Transport systems can be modelled in a uniform way as network systems. Planning and optimization in this context are typical examples of structured, combinatorial problems, such as scheduling, network flows, shortest paths and routing problems. However, the conditions and circumstances are induced by real-world demand. Therefore, a first step consists of transforming such complex practical problems into simplified models still describing their most important characteristics. After a model has been established, efficient methods and techniques are studied to develop solutions. Although mathematical solution methods and theoretically well-studied algorithms are available, the integration of many different optimization criteria and constraints calls for their talented and sophisticated use. Moreover, a combination of such mathematical methods with experiences from algorithms engineering and with extensive experimental studies, and of course new ideas, are essential.

Examples of research topics to be discussed are:

Timetable Construction
Although nice theoretical models for automatic timetable construction are found in the scientific literature, railway companies still mainly rely on non-automatic or at most semi-automatic systems for timetabling. There are few exceptions like the timetable construction system CADANS developed at CWI in cooperation with Railned and NS Reizigers. This package may serve as an example to other railway companies regarding its scientific content and innovative character in general. A long-term goal could be automatic timetabling for several European countries across national boundaries or even different public transport systems.

Timetable Information
Timetable information is a typical example of an optimization problem to be solved not only in national but also in a European context and not only for railway systems but for public transport in general. At the current state, the timetable information system Hafas, developed by the German company HaCon and provided by DB Systems is used by most European countries. Due to the huge size of the underlying data set, which is still

increasing, the improvement of the algorithmic kernel of such a system is still a challenge. Moreover, providing dynamic timetable information is still a great challenge.

Network Planning and Evaluation
National railway companies are starting to apply algorithms and methods from combinatorial optimization in order to evaluate the quality of their network respectively of planned extensions of it. Relevant aspects in this area are the facilitation of the projected growth of the demand in the near future and the trade-off between customer service and operational costs. One research direction is the study of the applicability and the relevance of network flow methods for analyzing these problems. A long term goal is the development of evaluation methods for the European railway network across national boundaries.

Dynamic Systems
An interesting and challenging issue is the dynamic setting of algorithmic problems arising in transport systems. For all research topics in transport systems, time is an important aspect, because modelling parameters and constraints vary over the time. Examples are dynamic aspects of network planning and evaluation or delay management. This subject also includes automatic reactions to distortions, a problem that may be studied as a distributed combinatorial optimization problem.

1.7 Cache-Oblivious and Cache-Aware Algorithms

Seminar No. **04301** Date **18.07.–23.07.2004**
Organizers: L. Arge, M.A. Bender, E.D. Demaine, Ch.E. Leiserson, K. Mehlhorn

A recent trend in algorithmic design and analysis is to pay increasing attention to the memory hierarchy of a modern computer, motivated by the major impact it has on the performance of algorithms. Many approaches and models have been developed, one of the more successful of which is the body of work in *external-memory algorithms*, which effectively models a two-level memory hierarchy, traditionally main memory and disk. Recently (1999), the concept of *cache-oblivious algorithms* was proposed as a powerful theoretical model with potential for major practical impact, especially for cache efficiency. The idea is to hide any parameters of the memory hierarchy – such as block transfer sizes and the size of each memory level – from the algorithm. This simple idea has powerful ramifications:

1. Cache-oblivious algorithms automatically adapt to arbitrary memory hierarchies.
2. Cache-oblivious algorithms can be analyzed on a simple two-level memory hierarchy, and then automatically perform as well on a complex multilevel memory hierarchy with particular page replacement strategies, limited associativity, etc.

Motivated by these exciting consequences, an increasing number of researchers have started to develop algorithms and data structures in this model. In the past 3 years, there have

been over a dozen papers developing efficient cache-oblivious algorithms and data structures. The field has been shown to have substantial depth and has led to new general techniques for maintaining data locality in a memory hierarchy.

Yet the field of cache-oblivious algorithms is still in its infancy. Several theoretical issues remain unsolved and ripe for exploration, bridging the gap between the best algorithms known in the external-memory and cache-oblivious contexts. On the practical side, the main issue is to what extent cache-oblivious algorithms are useful in the real world and throughout computer science. One issue here is how well the cache-oblivious model matches real caches, for example, in the context of the Translation Lookaside Buffer (TLB). In addition to the need for more thorough algorithmic experiments, the algorithms community needs to become aware of more practical problems for which cache-oblivious algorithms may be useful.

The goal of this seminar is to bring together people from various disciplines in order to address these issues and mature the field of cache-oblivious and other cache-efficient algorithms. These disciplines include both the algorithms community – specifically, external-memory algorithms, data structures, and experimental algorithmics, in addition to cache-oblivious algorithms – as well as various applied communities in computer science.

1.8 Algorithms and Complexity for Continuous Problems

Seminar No. **04401** Date **26.09.–01.10.2004**
Organizers: T. Müller-Gronbach, E. Novak, K. Petras, J.F. Traub

The goal of this workshop was to bring together researchers from different communities working on computational aspects of continuous problems.

Continuous computational problems arise in many areas of science and engineering. Examples include path and multivariate integration, function approximation, optimization, as well as differential, integral, and operator equations.

Understanding the complexity of such problems and constructing efficient algorithms is both important and challenging.

The workshop was of a very interdisciplinary nature with invitees from, e.g., computer science, numerical analysis, discrete, applied, and pure mathematics, physics, statistics, and scientific computation. Many of the lectures were presented by Ph.D. students.

Compared to earlier meetings, several very active research areas received more emphasis. These include quantum computing, complexity and tractability of high-dimensional problems, stochastic computation, and quantization, which was an entirely new field for this workshop.

Due to strong connections between the topics treated at this workshop many of the participants initiated new cooperations and research projects. The meeting was the eighth in a series of Dagstuhl-seminars on “Algorithms and Complexity for Continuous Problems”.

This topic belongs to the focus group research areas of the Working Group 1.1 of the International Federation for Information Processing. The work of the attendants was supported by a variety of funding agencies including the Deutsche Forschungsgemeinschaft, the National Science Foundation and the Defense Advanced Research Projects Agency (USA), the Australian Research Council, and the State Committee for Research (Poland).

Selected papers from the workshop will be published in a special issue of the Journal of Complexity.

The following topics were discussed:

1. Complexity and Regularization of Ill-Posed Problems
2. Non-Linear Approximation
3. Tractability of High-Dimensional Numerical Problems
4. Quasi-Monte Carlo Methods
5. Quantum Computing
6. Stochastic: Computation and Quantization
7. Global Optimization
8. Differential and Integral Equation

1.9 Algebraic Methods in Computational Complexity

Seminar No. **04421** Date **10.10.–15.10.2004**
Organizers: H. Buhrman, L. Fortnow, T. Thierauf

The seminar brought together researchers covering a wide spectrum of complexity theory. The focus on algebraic methods showed once again the great importance of algebraic techniques for theoretical computer science.

We saw a series of presentations by Andris Ambainis, Robert Spalek and Mario Szegedy. Ambainis described his improved method for showing lower bounds for quantum algorithms that provably beats the degree method. Spalek talked about his work with Szegedy showing that Ambainis techniques as well as different tools developed by Zhang, Laplante and Magniez, and Barnum, Saks and Szegedy all gave the same bounds. Szegedy, in his presentation, called this complexity measure sumPI and showed that the size of a Boolean formula computing a function f is at least $\mathsf{sumPI}^2(f)$. Further talks on quantum complexity considered lower bounds for formula size (Scott Aaronson), finite groups (Steve Fenner), and adversary bounds (Robert Spalek).

Discussions between Laplante, Lee and Szegedy at the workshop led to the recent announcement of even a stronger lower bound for Boolean formula complexity using a stronger version of sumPI complexity called maxPI.

Manindra Agrawal presented recent work of his students Neeraj Kayal and Nitin Saxena (the trio that showed a polynomial-time algorithm for primality testing) on rings given by a matrix describing the actions on the base elements. They show a randomized reduction from graph isomorphism to ring isomorphism and from factoring to #RI, counting the number of ring isomorphisms. They also show a polynomial-time algorithm for determining if there are any non-trivial automorphisms of a ring and that #RI is computable with an oracle for AM ∩ coAM. Agrawal conjectured that #RI is computable in polynomial time, a conjecture that would imply factoring and graph isomorphism have efficient algorithms.

In addition to Agrawal's presentation on ring isomorphism, we had a wide-range of talks on classical complexity. Lance Fortnow, building on work of Shaltiel and Umans, characterized the interesting question whether EXP is contained in NP with logarithmically bounded advice. Judy Goldsmith showed that the dominance problem for user preferences is PSPACE-complete. Various hypothesis in complexity theory were compared by John Hitchcock. Jörg Rothe located the Exact-Four-Colorability problem in the boolean hierarchy. Several probabilistic time classes were separated by Rahul Santhanam. Leen Torenvliet considered autoreducible sets and Falk Unger presented some new results on sparse self-reducible sets. Talks on circuit complexity were given by Anna Gal, Ryan O'Donnell and Denis Therien.

Troy Lee considered the symmetry of information in Kolmogorov complexity. Klaus Ambos-Spies presented algorithmic and resource-bounded genericity concepts. Tolerant property tester were investigated by Eldar Fischer, and automatic structures by Frank Stephan. Markus Maucher presented a very interesting relationship between the entropy of random source and the running time of (randomized) quicksort, where the random source is used for the choice of the pivot element. Rüdiger Reischuk considered string compression based on context-free grammars: the problem to determine the minimal size of a grammar that produces precisely one given string x was shown to be NP-hard for alphabets of larger size.

1.10 Practical Approaches to Multi-Objective Optimization

Seminar No. **04461** Date **07.11.–12.11.2004**
Organizers: J. Branke, D. Kalyanmoy, K. Miettinen, R.E. Steuer

As the name suggests, multi-criterion optimization involves optimization in the presence of more than one (conflicting) criteria. Multi-criterion optimization problems arise in a variety of real-world applications and the need for efficient and reliable solution methods is increasing. The main difference between single and multi-criterion optimization is that in case of the latter, there is usually no single optimal solution, but a set of equally good alternatives with different trade-offs, also known as Pareto-optimal solutions. In the absence of any other information, none of these solutions can be said to be better than the other. Usually, a decision maker is needed to provide additional preference information and to identify the most satisfactory solution. Depending on the paradigm

used, such knowledge may be introduced before, during, or after the optimization process. Multi-criterion optimization thus has to combine two aspects: optimization and decision support.

So far, there have existed basically two different approaches with corresponding fields of research: the classical approach and the evolutionary algorithm approach. The classical approach has a comparatively long history and subsumes a number of different algorithms, mostly developed by researchers from mathematics and operations research. These algorithms usually generate one solution and then adjust and re-optimize it according to the user's preferences. If they generate a set of alternatives, a small set of alternatives are usually generated sequentially. Many classical methods convert the problem with multiple criteria into one or a series of single criterion problems. On the other hand, evolutionary multi-objective optimization (EMO) is a relatively young research area, and is mostly grounded in computer science and engineering. Since evolutionary algorithms maintain a population of solutions throughout the optimization process, they are naturally suited to search for a large, representative set of Pareto-optimal solutions in parallel. In most cases, these solutions are created whithout the intervention of any decision-maker. Only after the optimization process has been completed, the decision maker chooses from the set of alternative solutions.

Both classical and evolutionary approaches have their merits and demerits, and both fields have become important and emerging topics in the area of informatics and operations research. But while both fields have resulted in numbers of efficient algorithms and successful applications published in dedicated books and major international journals and conferences, interactions among the fields are rarely found.

A major purpose of this seminar therefore was to bring together leading experts from both fields for discussions on the current state-of-the-art methodologies and common interests. The seminar was participated by 39 persons from 15 countries. It consisted of two tutorials, 27 talks, group discussions on three selected topics, a panel discussion, and software demonstrations.

Achievements of the Seminar / Feedback from Participants

The seminar was clearly unique because it brought together scholars from the two sides of the rapidly growing area of multi-criterion optimization. Prior to the seminar, people from the two sides, coming from different disciplines, hardly knew each other. However, as result of the seminar, it is clear that two sides are now considered as comprising the whole of the field of multi-criterion optimization in the most modern interpretation of the term.

The discussions were lively and certainly helped to clarify the terminology used in different groups, which should make reading and understanding each other's research paper easier. Active collaboration between the fields was fostered. As one example, some participants, from both classical and EMO fields even made a head-start to work on a joint project of outlining an interactive multi-criterion optimization method bringing together and hybridizing ideas utilized in the two fields.

At the closing session, the organizers collected feedback about the seminar, which is summarized in the following. All participants found the seminar very successful and were very excited about the idea of bringing together researchers from the two fields of multi-criterion optimization and expressed their interest in participating in such a seminar again. Participants appreciated the excellent opportunity to get to learn about the state-of-the-art, to discuss, to interact, to exchange ideas and to get new ideas. They liked the possibility of meeting many experts and asking them directly on their subjects of interests. It was great how people from classical and EMO fields were able to establish collaborative arrangements that would not have been possible without such a seminar. They also liked the idea of working groups and their results. It was emphasized that we should cooperate as much as possible including newsgroups, a paper repository, and website. Some participants even said that this seminar was the most fruitful meeting for a decade for them! Finally, the participants mentioned the Dagstuhl atmosphere as a really significant ingredient in the success of the seminar.

Chapter 2

Geometry, Image Processing, Graphics

2.1 Geometric Properties from Incomplete Data

Seminar No. **04131** Date **21.03.–26.03.2004**
Organizers: R. Klette, R. Kozera, L. Noakes, J. Weickert

Computer vision and image analysis requires interdisciplinary collaboration between mathematics and engineering, especially in the area of high-accuracy measurements of length, area, curvature, motion parameters and other geometrical quantities from acquired image data. The seminar will bring together researchers in computer vision, engineering and mathematics who are working in this area.

Recent proceedings of international conferences on computer vision emphasize mathematical methods for measurement problems, drawing on classical and differential geometry, mathematical analysis, optimization, topology, statistics, stochastic equations, differential equations and numerical analysis.

The seminar will focus on interdisciplinary work on estimation of curves, surfaces and motion from image data that is mathematically incomplete, either through digitization, contamination by noise, or some other indirectness in the measurement process. Problems include

1. estimation of lengths of curves and areas of surfaces from digitized images (geometric feature analysis)
2. estimation of curves and surfaces from samples of unparameterized data (shape recovery)
3. estimation of objects from incomplete volume data such as X-ray, MRI, CT scans (model-based shape extraction)
4. estimation of landmarks on solid objects from multiple camera images (structure from motion)

5. estimation of rigid body motion, particularly in connection with computer vision.

Participants will be invited to discuss these tasks and relationships with mathematics, especially approximation theory, numerical analysis and differential geometry. Contributions in digital topology, computational geometry and complexity, geometric modelling, optimization and differential equations are also in the scope.

As a specific example, recent work of Katoulokis on space-carving has attracted a lot of interest, including attempts to speed up the process. One of the steps, to quickly estimate the photo-hull, can be viewed as a nonstandard problem in approximation theory, where unparameterized direction data is available from unknown points on a curve or surface.

All seminar participants will be expected to contribute short (15-20 minutes) or long talks (30-45 minutes). Refereed seminar papers will be published in a research monograph volume in Springer Lecture Notes in Computer Science, with submission following the seminar. The seminar schedule will be characterised by flexibility, working groups, and sufficient time for focused discussions. Working groups will be formed on day one of the seminar for smaller and more focuses meetings during the week.

2.2 Perspectives Workshop: Visualization and Image Processing of Tensor Fields

Seminar No. **04172** Date **18.04.–23.04.2004**
Organizers: H. Hagen, J. Weickert

Motivation

Recently, matrix-valued data sets (so-called tensor fields) have gained significant importance in the fields of scientific visualization and image processing. This has been triggered by the following developments:

- Novel medical imaging techniques such as diffusion tensor magnetic resonance imaging (DT-MRI) have been introduced. DT-MRI is a 3-D imaging method that yields a diffusion tensor in each voxel. This diffusion tensor describes the diffusive behaviour of water moluecules in the tissue. It can be represented by a positive semidefinite 3 x 3 matrix in each voxel.

- Tensors have shown their use as a general tool in image analysis, segmentation and grouping. This also includes widespread applications of the so-called structure tensor in fields ranging from motion analysis to texture segmentation.

- A number of scientific applications require to visualize tensor fields. The tensor concept is a common physical description of anisotropic behaviour, especially in solid mechanics and civil engineering (e.g. stress-strain relationships, inertia tensors, diffusion tensors, permittivity tensors).

Problems

This has led to a number of challenging scientific questions, e.g.

- How can one visualize these high-dimensional data in an appropriate way?
- What are the relevant features to be processed? Is it better to have component-wise processing, to introduce some coupling between the tensor channels or to decompose the tensors in their eigenvalues and eigenvectors and process these entities separately?
- How should one process these data such that essential properties of the tensor fields are not sacrificed? For instance, often one knows that the tensor field is positive semidefinite. In this case it would be very problematic if an image processing method would create matrices with negative eigenvalues.
- How can structure of tensor fields be addressed? Topological methods have been used in visualization, but more research in this area is needed.
- How should one adapt the processing to a task at hand, e.g. the enhancement of fibre-like structures in brain imaging? This may be very important for a number of medical applications such as connectivity studies.
- How can one perform higher-level operations on these data, e.g. segment tensor fields? Current segmentation methods have been designed for scalar- or vector-valued data, and it is not clear if and how they can be extended to tensor fields.
- How can one perform operations on tensor fields in an algorithmically efficient manner? Many tensor fields use 3 x 3 matrices as functions on a three-dimensional image domain. This may involve a very large amount of data such that a clear need for highly efficient algorithms arises.

Since this research area is very young, these fundamental questions have not been solved yet. In image processing, e.g., the filtering of tensor fields has not been investigated before 2000. Moreover, a lack of interdisciplinary interaction is another reason for the numerous unsolved problems in this area: Many medical imaging people are unaware of recent progress in the tensor-based image analysis area, while image processing specialists do not know much about recent medical imaging techniques such as DT-MRI. Their research would also benefit significantly if they had the possibility to visualize the results of their tensor-based filters in a suitable way. On the other hand, most computer graphics specialists do not yet use advanced image processing methods to smooth and enhance their tensor fields.

Goals

In this Dagstuhl seminar, we want to bring together experts from scientific visualization, image processing and medical imaging in a real interdisciplinary workshop. Each invited

participant has contributed to specific aspects in the area of tensor field imaging. The interaction of these participants will clarify the needs that every group has and the scientific perpectives it can contribute to the field of tensor imaging. We expect that identifying these mid-term perspectives should lead to a significant boost of the scientific output on tensor-based imaging.

2.3 Imaging Beyond the Pin-hole Camera. 12th Seminar on Theoretical Foundations of Computer Vision

Seminar No. **04251** Date **13.06.–18.06.2004**
Organizers: K. Daniilidis, R. Klette, A. Leonardis

The world's first photograph was taken by Joseph Nicphore Niépce (1775-1833) in 1826 on his country estate near Chalon-sur-Sane, France. The photo shows parts of farm buildings and some sky. Exposure time was eight hours. Nipce used a pinhole camera, known as camera obscura, and utilized pewter plates as the support medium for the photographic process. The camera obscura, the basic projection model of pinhole cameras, was first reported by the Chinese philosopher Mo-Ti (5th century BC): light rays passing through a pinhole into a darkened room create an upside-down image of the outside world.

Cameras used since Niépce are basically following the pinhole camera principle. The quality of projected images improved due to progress in optical lenses and silver-based film, the latter one replaced today by digital technologies. Pinhole-type cameras are still the dominating brands, and also used in computer vision for understanding 3D scenes based on captured images or videos.

However, different applications have pushed for designing alternative architectures of cameras. For example, in photogrammetry cameras are installed in planes or satellites, and a continues stream of image data can also be created by capturing images just line by line, one line at a time. As a second example, robots require to understand a scenery in full 360° to be able to react on obstacles or events; a camera looking upward into a parabolic or hyperbolic mirror allows this type of omnidirectional viewing. The development of alternative camera architectures also requires to understand related projective geometries for the purpose of camera calibration, binocular stereo, or static or dynamic scene understanding.

This abstract collection reports about contributions given at a seminar at the international computer science center in Dagstuhl (Germany) addressing basics and applications of alternative camera technologies, in particular in the context of computer vision, computer graphics, visualisation centers, camera producers, or application areas such remote sensing, surveillance, ambient intelligence, satellite or super-high resolution imaging. Examples of subjects are geometry and image processing on plenoptic modalities, multiperspective image acquisition, panoramic imaging, plenoptic sampling and editing, new camera technologies and related theoretical issues.

This abstract collection is structured into five parts on (1) sensor geometry for different camera architectures, also adressing calibration, (2) applications of non-pinhole cameras for analyzing motion, (3) mapping of 3D scenes into 3D models, (4) navigation of robots using new camera technologies, and (5) on specialized aspects of new sensors and other modalities.

New results and specific research strategies have been discussed at this seminar to approach this highly complex field. The seminar intention was to discuss theoretical fundamentals related to those issues and to specify open problems and major directions of further development in the field of new camera technologies related to computer vision, computer graphics and related applications. The seminar schedule was characterised by flexibility, working groups, and sufficient time for focused discussions.

There will be an edited volume of seminar papers (within the Kluwer series) with an expected publication date in early 2006.

Chapter 3

Artificial Intelligence, Computer Linguistic

3.1 Evaluating Embodied Conversational Agents

Seminar No. **04121** Date **14.03.–19.03.2004**
Organizers: Z. Ruttkay, E. André, K. Höök, W. Lewis Johnson, C. Pelachaud

In the past years we have seen an increasing number of ECAs (Embodied Conversational Agents), as results (and sometimes also as medium) of academic research as well as in commercial applications in fields like information providing (news reader, info kiosk), user interfaces for deaf (lip-reading, signing avatars), education, commerce and entertainment. Due to the novelty of the field and the complexity of the problems, different paradigms and tools have been used to produce the individual ECAs, often tailored to the special needs of an application or of a research issue.

The comparison and re-use of ECAs has not been dealt with to a sufficient degree, in spite of the high costs of developing yet new ECAs. There have been studies that indicate added value of particular features of ECAs, but it remains to be determined in what situations fully functional ECAs are advantageous, and for what purpose. The sporadic evaluation experiments concentrated on special issues and under restricted circumstances. All the same, several of these partial evaluations have already produced surprising and for the future development and application of ECAs essential results, e.g. on the different preferences of different users.

For the evaluation and design guidelines of ECAs one should rely on results and methodologies of related fields like HCI, psychology, cultural anthropology and even arts. The different languages and methodologies of these disciplines and lack of common forums have hindered the cross-fertilization. Feedback from the industry on commercial applications would also be very useful.

Motivated by the above observations, the main objectives of the seminar were:

- to review the results, tools and resources;

- identify key problems and future research directions;
- initiate further activities to trigger and dissimilate work on ECA design and evaluation.

3.2 Logic Based Information Agents

Seminar No. **04171** Date **18.04.–23.04.2004**
Organizers: J. Dix, Th. Eiter, E. Franconi

Logic-based methods have a great potential as a toolbox for the development of information technology infrastructure which is able to provide and handle advanced services. With the creation of the world wide web, and the advent of the internet as a communication backbone for connecting people, more and more data and information is becoming available; however, our current methods for managing this wealth of information, including searching, integrating, and updating, are still at an early stage. Quite some efforts will be needed to research the foundations of this as well practical methods. This concerns particular methods for well-defined information processing tasks, as well as the design of systems for intelligent information processing. Due to the distributed nature of the web, and the local autonomy of information sites, this requires in particular the design and development of societies of information agents, which need to cooperate for providing information services as desired by the end user.

The Dagstuhl Seminar 04171 (Logic Based Information Agents) brought together researchers and developers who are involved in the research for methods and the design of systems for agent-based information processing, and in particular for searching, fusing, and cleaning distributed data and knowledge, using tools and methods from computational logic.

The seminar consisted of 20 presentations, which can be classified into the following subgroups:

1. Semantic Web and Description Logics;
2. P2P Data Integration;
3. Agent Systems—agent architectures, agent societies, and logical models of agents.

An open discussion was held on Friday morning about different opinions among participants on the following questions:

1. Should P2P data integration approaches care about agent theories?
2. Conversely, is P2P data integration just “simple” agent theory?
3. What is the use of agent theories in the semantic web? Or, vice versa, what can agent theories gain from semantic web methods?

4. What are "missing links" between (1) semantic web, (2) data integration, and (3) agent systems?
5. How could the current state of the art in (1), (2), and (3) be combined into a uniform framework? Is this desired?
6. Have you seen something from (1), (2) or (3) that might be useful in your field of expertise?

It is clear from the discussion that the fields of (1) semantic web, (2) P2P data integration, and (3) agent systems are closely connected. On one hand, the semantic web technology requires involvement and efforts from agent theories, and on the other hand, semantic web provides a ground for both P2P data integration and agent systems, which can be built on top of it. With more complex P2P systems, more agent technologies would be required and helpful, and agent technologies would play more important role.

In summary, the seminar provided a good opportunity for closer cooperation between the researchers from the areas of agent systems, semantic web, and P2P data integration. It helped to get a better understanding of the ways in which logic-based information agent systems may be built in the future, and helped to develop more efficient implementations of such systems.

Chapter 4

Programming Languages, Compiler

4.1 Dependently Typed Programming

Seminar No. **04381** Date **12.09.–17.09.2004**
Organizers: T. Altenkirch, M. Hofmann, J. Hughes

The Dagstuhl seminar (04381) on Dependently Typed Programming brought together researchers from all over the world who are interested in the use of dependent types in programming. An emerging topic was the interaction of the functional programming community and the Types community: an example is the use of GADTs in Haskell, which represent a restricted use of dependent types in Haskell, while on the other hand proof systems like COQ allow the expression of many functional programming idioms. Emerging languages and systems, like Epigram, attempt to unify functional programming and Type Theory based proof development environments. Discussions during the seminar centred on the question how to integrate dependent types in real programming languages and on the pragmatic and theoretical questions raised by doing this.

4.2 Synchronous Programming – SYNCHRON'04

Seminar No. **04491** Date **28.11.–03.12.2004**
Organizers: S.A. Edwards, N. Halbwachs, R. v. Hanxleden, T. Stauner

The goal of the seminar was to bring together researchers and practitioners of synchronous programming, and furthermore to reach out to relevant related areas and industrial users. With a record participation in this year's SYNCHRON workshop and a broad range of topics discussed, the aims seem to have been well-met. The program of the seminar was composed of around thirty presentations, all of which included extensive technical discussions. The fields covered included synchronous semantics, modeling languages, verification, heterogeneous and distributed systems, hardware/software integration, reactive processing, timing analyses, application experience reports, and industrial requirements.

Particularly successful this year were presentations from the automotive industry. Stefan-Alexander Schneider and Thomas Stauner both discussed issues with real-time software development at BMW. Matthias Hoffmann represented DaimlerChrysler.

Synchronous Languages

Historically, the first synchronous language is Esterel, developped at the Centre de Mathématiques Appliquées (CMA) of Ecole des Mines de Paris, in Sophia-Antipolis, France, and later joined by people from INRIA. It is an imperative language that was originally inspired by CCS and SCCS. Esterel introduces constructs like preemption and communication by synchronous broadcast. It is devoted to the programming of discrete event systems. Esterel Technologies now markets an industrial version of the Esterel compiler. There exist several other synchronous languages. This is just a selection, presented in chronological order:

- Lustre is a data-flow declarative functional language also inspired by Lucid. The Scade tool, initially developed by Verilog and Aerospatiale is based on Lustre. Scade is now marketed by Esterel Technologies.

- Signal is also a data-flow declarative language, but it is relational instead of functional like Lustre. In this sense, it is more general than Lustre. Polychrony is the public domain Signal compiler, while Sildex is the commercial tool developed by TNI-Valiosys.

- Argos is a purely synchronous version of the well known Statecharts formalism, which yields a number of advantages. In particular, Argos has a compositional semantics. SyncCharts and Mode Automata are both inspired from Argos.

- Polis is a graphical tool for implementing Codesign Finite State Machines (CFSM). The model of computation behind CFSMs is a set of synchronous FSMs communicating asynchronously; It is therefore known as Globally Asynchronous Locally Synchronous (GALS). The Cierto VCC tool developed by Cadence is based on Polis.

- SL, the Synchronous Language, is a variant of Esterel where hypotheses about signal presence or absence are not allowed. Whether a given signal is present or absent can only be decided at the end of a synchronous instant, hence reaction to a signal is delayed until the next instant. The main advantage is that causality problems are avoided. SL was the starting point of many other synchronous languages such as Sugar Cubes and Junior.

While Esterel, Argos, and SL are more suited to discrete event systems, Lustre, Signal and Polis are very close to the specification formalisms used by automatic control engineers: block diagrams, differential equations, data flow networks, automata, and so on.

Industrial Impact

Synchronous languages have recently seen a tremendous interest from leading companies developing automatic control software for critical applications, such as Schneider, Dassault, Aerospatiale, Snecma, Cadence, Texas, and Thomson. For instance, Lustre is used to develop the control software for nuclear plants and Airbus planes. Esterel is used to develop DSP chips for mobile phones, to design and verify DVD chips, and to program the flight control software of Rafale fighters. And Signal is used to develop digital controllers for airplane engines. The key advantage pointed by these companies is that the synchronous approach has a rigorous mathematical semantics which allows the programmers to develop critical software faster and better.

In summary, synchronous programming is an interesting approach for designing and programming automatic control software. Synchronous languages have a well-founded mathematical semantics that allow ideal temporal constructs as well as formal verification of the programs and automatic code generation. We believe they are ideally suited to programming automatic control software because they are close to the classic specification formalisms used by control engineers, and also because they offer code generation tools that avoid the tedious and error-prone task of implementing the control algorithm after having specified it. These nice features have been confirmed by their recent successes in the automatic control industry.

Chapter 5

Software Technology

5.1 Perspectives Workshop: Empirical Theory and the Science of Software Engineering

Seminar No. **04051** Date **25.01.–29.01.2004**
Organizers: J.D. Herbsleb, W.F. Tichy

Software engineering has increasingly come to depend on empirical studies to validate claimed contributions to the field, yet the value of empirical studies is still a source of some controversy. It is very difficult to summarize and organize the knowledge gained from experiments, to clearly state what is known, to understand where the frontier of knowledge is, and to identify the hypotheses that are currently most in need of empirical test. In the mature sciences, these are roles played by scientific theory.

With precious few exceptions, scientific theory (as opposed to mathematical theory) is in short supply in software engineering. Fostering the development of insightful, testable theory is critical to the future of scientific progress, cumulative knowledge, and systematic investigation in software engineering. In this seminar, we will bring together the most thoughtful empirical researchers in software engineering, along with a few scientists from related disciplines with theoretical maturity, such as economics, information systems, sociology, anthropology, and psychology, to 1) discuss the nature of theory in software engineering, and 2) generate theoretical ideas.

5.2 Language Engineering for Model-Driven Software Development

Seminar No. **04101** Date **29.02.–05.03.2004**
Organizers: J. Bézivin, R. Heckel

Model-driven approaches to software development require precise definitions and tool support for modeling languages, their syntax and semantics, their notions of consistency

and refinement, as well as their mappings to the implementation level. In order to support model-driven development in a variety of contexts, we must find efficient ways of designing languages, accepting that definitions are evolving and that tools need to be delivered in a timely fashion.

In this respect, language definitions are not unlike software. Thus, a discipline of **language engineering** is required to support the design, implementation, and validation of modeling languages with the goal to deliver languages at low cost and with high quality.

An important contribution of any engineering science, besides the actual technology provided, is the meta knowledge about what are the relevant concerns to be addressed, what are the possible solutions, and what concern is best addressed in a given context by which kind of technology.

It is understood that different concerns of language engineering, like the definition of abstract syntax and well-formedness rules, operational and denotational semantics, consistency and refinement relations, and model transformations, will, in general, require technologies from different domains.

A framework for classifying, choosing, and relating different solutions domains is provided by the concept of technological spaces. A technological space is a working context with a set of associated concepts, body of knowledge, tools, acquired skills and possibilities, often associated to a given community. Well-known examples include XML, UML meta modeling, graph transformation, algebra and logic, programming languages, etc.

It has been the goal of the seminar to investigate relevant concerns and promising solution domains for language engineering, learn from specific solutions presented by the participants, and attempt a provisional classification and mapping. To illustrate problems and available solutions, a sample language engineering problem was proposed and elaborated.

5.3 Perspectives of Model-Based Testing

Seminar No. **04371** Date **05.09.–10.09.2004**
Organizers: E. Brinksma, W. Grieskamp, J. Tretmans, E. Weyuker

Software testing

Software invades everywhere in our society and life, and we are increasingly dependent on it. This applies to all kinds of software: software in safety critical systems such as airplanes, in consumer products, in mobile phones, in telecom switches, in pace makers, in process control systems, in financial systems, in administration systems, etc. Consequently, the quality of software is an issue of increasing importance and growing concern.

Systematic testing is one of the most important and widely used techniques to check the quality of software. Testing, however, is often a manual and laborious process without effective automation, which makes it error-prone, time consuming, and very costly. Estimates are that testing consumes between 30-50% of the total software development costs.

Moreover, the majority of the testing activities take place in the most critical phase of software development, viz. at the end of the project just before software delivery.

The tendency is that the effort spent on testing is still increasing due to the continuing quest for better software quality, and the ever growing size and complexity of systems. The situation is aggravated by the fact that the complexity of testing tends to grow faster than the complexity of the systems being tested, in the worst case even exponentially. Whereas development and construction methods for software allow the building of ever larger and more complex systems, there is a real danger that testing methods cannot keep pace with construction, so that these new systems cannot sufficiently fast and thoroughly be tested anymore. This may seriously hamper the development of future generations of software systems.

Model based testing

One of the new technologies to meet the challenges imposed on software testing is *model-based testing*. In model-based testing a *model* of the desired behaviour of the *system under test* (SUT) is the starting point for testing. Model-based testing has recently gained attention with the popularization of modeling itself both in academia and in industry. The main virtue of model-based testing is that it allows test automation that goes well beyond the mere automatic execution of manually crafted test cases. It allows for the algorithmic generation of large amounts of test cases, including test oracles, completely automatically from the model of required behaviour. If this model is valid, i.e. expresses precisely what the system under test should do, all these tests are also provably valid. Moreover, these models can, in principle, also be used for defining e.g. specification coverage metrics and test selection with mathematical rigour, so that quantifiable confidence is obtained, that a product faithfully conforms to its specification.

From an industrial perspective, model-based testing is a promising technique to improve the quality and effectiveness of testing, and to reduce its cost. The current state of practice is that test automation mainly concentrates on the automatic execution of tests. For this, a multitude of commercial test execution tools is available, but these tools do not address the problem of test generation. Model-based testing aims at automatically generating high-quality test suites from models, thus complementing automatic test execution.

From an academic perspective, model-based testing is a natural extension of formal methods and verification techniques, where many of the formal techniques can be reused. Formal verification and model-based testing serve complementary goals. Formal verification intends to show that a system has some desired properties by proving that a model of that system satisfies these properties. Thus, any verification is only as good as the validity of the model on which it is based. Model-based testing starts with a (verified) model, and then intends to show that the real, physical implementation of the system behaves in compliance with this model. Due to the inherent limitations of testing, such as the limited number of tests that can be performed, testing can never be complete: testing can only show the presence of errors, not their absence.

The interest in model-based testing from both industry and academia provides perspectives for academic-industrial cooperation in this area. This is also reflected in the relatively high

industrial participation in the seminar, with researchers from Siemens, DaimlerChrysler, IBM, France Telecom, and Microsoft attending, and even co-organizing.

The Seminar

The aim of the seminar *Perspectives of Model-Based Testing* was to bring together researchers and practitioners from industry and academia to discuss the state of the art in theory, methods, tools, applications, and industrialization of model-based testing, and to identify the important open issues and challenges.

In the past an analogous seminar had been organized: *Test Automation for Reactive Systems - Theory and Practice* (Ed Brinksma, Jan Peleska and Michael Siegel, Dagstuhl seminar 98361; Report 223; September 1998). Although at that time there were already quite a few research groups active in the area, it was only the beginning of developments in model-based testing. Significant progress has been made since then, and we had the impression that another Dagstuhl seminar on the topic was justified.

The presentations and discussions in the seminar addressed a broad spectrum of topics, which together gave a good overview of the current state of the art, the perspectives, and the open questions of model-based testing (MBT). Views from both academic and industrial perspectives were presented, different kinds of modelling and specification formalisms were used, and different test generation techniques, new ones as well as extensions of existing ones, were discussed. We present some of our observations.

- A couple of presentations came from industrial application domains: embedded automotive software, avionics, service platforms, telecom software, and general office software. It seems that at the moment the area of embedded and technical software is the most fertile domain for MBT, with Microsoft being the most notable exception. At Microsoft, applications of MBT go also beyond embedded and technical systems into the domain of general application programming interfaces and user interfaces.

  The industrial presentations stressed that testing is an important aspect of software development, that automated testing is a necessity, and that MBT has a role to play. Moreover, they identified a couple of challenges and open issues:

  - It is important for MBT to deal the imperfect and incomplete real world, in which requirements are never complete, and specifications are always partial or loose.
  - The development of a software product is not a self-contained, one shot activity, but a software product evolves: it is iteratively developed in increments, it comes in different versions, configurations, and releases, and it is combined with other software products and components introducing all kind of compatibility problems. MBT should be able to cope with these issues.
  - An important question is how to obtain the formal models for MBT, and how to combine application domain knowledge and modelling knowledge. This issue is aggravated by the fact that there is no consensus yet about a suitable language for expressing these models.

  - An issue for many new techniques, and for MBT in particular, is the question of scalability.
  - To be successful, MBT must be integrated in the development process.
  - Not only detecting errors is important, but also locating and diagnosing errors. Combining MBT with model-based diagnosis may be advantageous.

- Techniques to support MBT are drawn from many different areas like model checking, control and data flow analysis, static analysis, abstract interpretation, theorem proving, constraint solving, and run-time verification. These techniques are combined with traditional approaches to testing such as equivalence partitioning.

- The main emphasis during this seminar was on model-based testing of functionality. Other approaches to testing concerned user-profile based testing, where techniques like Markov-chain usage models and scenario-based statistical testing come into play.

- Whereas in the 1998-seminar a couple of basic, completely new test generation methods were presented, the current seminar had more emphasis on the investigation of extensions of these basic test generation methods, and on combinations of different methods and techniques. In particular, extensions for real-time testing, and combinations of state-based testing and data-oriented testing were elaborately discussed. Many of the presented techniques were in some sense "symbolic": symbolic test case generation, symbolic **ioco**, symbolic execution, symbolic transition systems, etc. Also extensions towards asynchronous testing, queue-based testing, and testing with action refinement were presented.

- Different kinds of languages and formalisms are used as the basis for MBT, such as FSM (Finite State Machines, or Mealy machines), different versions of labelled transition systems, (primitive recursive) functions, process algebra (e.g., CSP + CASL), ADT (Abstract Data Types), MSC (Message Sequence Charts), Spec#, etc. Most of them can be classified as formal, i.e., have a formal syntax and semantics, but also the less formal but industrially more often used notation UML was discussed as the basis for MBT. It was felt, however, that the lack of a precise semantics for UML, in particular of the dynamic behaviour part, hampers its use in model-based testing.

- Important, unsolved questions are how to select test cases, how to measure the quality of the selection, when to stop testing, how to quantify the remaining risk, and how to draw conclusions about the quality of a tested product. These problems are even more prominent in MBT, since the quantity of generated test cases is almost unlimited. These issues, however, were not often addressed during seminar, which might be due to lack of progress on these important issues.

- To compare different MBT methods and tools it would be nice if there were a set of benchmarks, for example, sets of specifications, models, implementations, and mutants, to which different tools and methods could be applied.

- Controversy appeared on the question whether the average user should be aware of the complexity of using formal methods in general, and MBT in particular, or

whether it should all be invisibly hidden in a tool, i.e., should MBT be push-button, or is this an infeasible dream?

Conclusion

The presentations at the seminar gave a good insight into what has been achieved in the area of model-based testing, and, even more important, they gave clear indications of what has to be done before we can expect widespread industrial use of model-based testing.

Compared with the 1998-seminar, the area of model-based testing has certainly matured, with expanding interest also from industry. The feasibility of model based testing has been shown, more groups are involved, more theories are supported by tools, some of them close to industrially strength tools, and there are successful applications. Software testing has become a respectable research area.

The prospects for model-based testing to improve the quality and to reduce the cost of software testing are positive, but still more effort is needed, both in developing new theories and in making the existing methods and theories applicable, e.g., by providing better tool support.

To inspire a possible research agenda in model-based testing, we concluded the seminar with a discussion of a list of the top 10 challenging issues in model-based testing:

1. Measures for coverage and test quality: Model-based test generation algorithms can produce many test cases, but there is no method yet to compare the quality of two test suites. Quantification of test quality is desirable to compare test suites and to select the best one.

2. Test purposes and test scenario control: Often, it is necessary to guide or control the model-based generation of test cases so that the interesting, error-prone, or tricky parts of an SUT are really tested, i.e, the test purpose is reached. How to identify and specify this purpose, and how to control and guide the generation of tests is still unclear.

3. Merging different models: Many test generation methods work for particular aspects of behaviour, e.g., state-based models mainly test for control flow. For real systems many aspects must be tested at the same time: state, control flow, data flow, data transformation real-time, etc. This requires integration of the corresponding modelling formalisms, and of the test generation methods.

4. The role of quiescence, timed and untimed: An SUT should do what it is required to do. Doing nothing is a particular form of such a requirement. Doing nothing is technically expressed as being *quiescent*. Quiescence is treated differently in different test generation methods. Better understanding of quiescence is necessary, in particular, when real-time is involved.

5. Modelling method invocations: Many test generation methods have their origins in a message-oriented paradigm. Current component-based systems work differently:

they are based on the object-oriented paradigm of method invocations. The modelling of method invocations for testing is not yet well-understood, in particular if parallelism and multi-threading is involved, so that several method invocations can exist concurrently.

6. Integration of techniques: The boundaries between such techniques as model-based testing, model checking, static analysis, abstract interpretation, theorem proving, constraint solving, run-time verification, etc. diminish. Integration of these techniques is necessary to be able to choose for every task the best combination of techniques.

7. Modelling test interfaces: For the execution of tests, the tester is connected to the SUT via some kind of test interface. The behaviour of this interface, e.g., an operating system *pipe* which behaves as a *FIFO queue*, must be taken into account when tests are generated. Research on different kinds of test interfaces and their influence on test generation and observation is desirable.

8. Tool architecture frameworks: Integration and interoperability of different (test) tools is desirable, e.g., interoperability of test generation tools and (on-the-fly) test execution tools.

9. Model based testing of non-functional properties: Most theory, methods, and tools for model-based testing have been devoted to testing of functionality. Model-based testing of other quality characteristics, such as security, reliability, performance, usability, etc., often referred to as non-functional properties, is an interesting field of research.

10. Promoting MBT in industry: To advance the industrial usage of MBT, it is necessary that the methods scale well to industrially sized problems, that they are sold with the right level of expectation, and that feedback from case studies is used in the next generation of MBT methods and tools.

The general opinion was that seminar was successful, and it was felt that, apart from the usual publication and presentation fora (e.g., FATES, ISSTA, MBT, TACAS, CAV, TestCom, ...) another Dagstuhl meeting on model-based testing should be organized in the future.

5.4 Architecting Systems with Trustworthy Components

Seminar No. **04511** Date **12.12.–17.12.2004**
Organizers: R. Reussner, J. Stafford, C. Szyperski

Component software technologies attract much attention for their promise to enable scaling of our software industry to new levels of flexibility, diversity, and cost efficiency.

Yet, these hopes collide with the reality that assemblies typically suffer from the proverbial "weakest link" phenomenon. If a component is used in a new compositional variation, then it will likely be stressed in a new way. Asserting useful properties of assemblies based on the used composition schema and theory requires a firm handle on the properties of the components being composed. For such assertions to hold, components need to meet their advertised properties, even if used under circumstances not explicitly envisaged by their developers. A component that fails to do so becomes a weak link of its hosting assembly and may cause the entire assembly to not meet its advertised properties.

In contrast, components that promise to be a strong link in their assemblies can be called 'trustworthy' and ways to get to the construction and proper use of such components are the subject of this seminar. Transitively, the seminar is also after trustworthy assemblies: assemblies that reliably meet their requirements based on trustworthy components and solid composition methods.

None of the weakest link phenomenon is a new observation, but the recent trend to move to dynamic and late composition of non-trivial components exasperates the problem. A concrete example promising deep wide-spread relevance are web services. The problem space is complex and multi-faceted. Practical solutions will have to draw on combined insights from a diverse range of disciplines, including component software technology, software engineering, software architecture, dependable systems, formal methods, as well as areas such as type systems and proof-carrying code.

A lot of good and sometimes even groundbreaking work has been performed in the focus area of this seminar, but much remains open. Bringing together many of the key minds in the various contributing areas to engage in this week-long seminar of mingling and discussions promises to spark some new key ideas and insights, ideally leading to new collaborative efforts.

To spark discussions, the seminar organizers propose a small set of core problems:

- measurement and normalization of non-functional properties,
- modular reasoning over non-functional properties,
- capture of component requirements in interfaces and protocols
- interference and synergy of top-down and bottom-up aspects,
- duality of componentization and architecture,
- system properties (non deadlocks, liveness, fairness, etc.)
- opportunities for correctness by construction/static checking

All of these are considered hard today and yet, all of them, if solved appropriately, promise the creation of key stepping stones towards an overall approach yielding trustworthy components as well as trustworthy compositions. It is likely that any such approach supports a multitude of more specialized disciplines and methods, targeting different requirement

profiles at the assembly level. Examples would include cases that require tight resource management or real-time characteristics.

Outcomes of the seminar will likely shape closer characterizations or answers to questions such as:

- Depending on the system-property to reason about, what are suitable techniques, and
- what component interface information do they require?
- Where are principal limitations of reasoning over a given system-property (depending on the reasoning technique)?
- Do certain system-properties conflict (e.g., performance - security)? For those pairs of conflicting properties, how can one find tradeoffs systematically?

Chapter 6

Applications, Multi-Domain Work

6.1 Atomicity in System Design and Execution

Seminar No. **04181** Date **25.04.–30.04.2004**
Organizers: C. Jones, D. Lomet, A. Romanovsky, G. Weikum

The Dagstuhl Seminar 04181 (Atomicity in System Design and Execution) brought together researchers and industrial practitioners from four fields: formal methods, fault tolerance, databases, and architecture to discuss the notion of Atomicity and its pervasive and effective use in all of their disciplines. The underlying purpose of Atomicity is to produce highly dependable, well architected, understandable, and verifiable computer systems or components. Members from these four fields engaged in an extremely productive discussion of ideas, techniques, and mindsets that is very difficult to achieve in a conference setting, where the scope is usually much narrower. In particular, the recent Dagstuhl Seminar provided ample opportunity for presenting seminal ideas, introducing new research, and engaging in intensive questioning and dialog. In short, it was a unique opportunity to learn from fields other than one's narrower specialty. We are writing an "Atomicity manifesto", recapping our Dagstuhl discussions and the insights gathered from across multiple fields, with the aim of encouraging by example more cross-discipline interactions.

The unifying framework for that seminar was the notion of "Atomicity" which cuts across many areas of Computing Science. It can refer to both core proprieties of system execution and design techniques. Atomicity is fundamental in dealing with the complexity of modern systems. It is introduced to simplify reasoning about and dealing with concurrency and fault tolerance. Atomicity is frequently introduced with some form of system structuring to provide a basis for encapsulating complex system parts. Atomicity is a ubiquitous concept in computer science and has been studied from many different angles.

The consensus among participants was that Seminar 04181 had been a very tasty first course in stimulating interactions between communities. The flexible plan of the first seminar initiated some real understanding but there was also a desire to move on to the next course. For example: discussions about how and when to use formal methods did not proceed far enough for us to understand their appropriate role and timing; it was unclear how database approaches can be applied outside of database systems; we speculated about

hardware and programming language support for Atomicity without understanding fully how to provide or how to use such "built-in" support; we discussed the role of consensus and atomic broadcast in some settings, but did not explore this far enough.

We believe that our research communities would greatly benefit by coming together again to further discuss how the various perspectives and technologies involving Atomicity might be applied in practice and embedded in a more comprehensive theoretical framework. Our intent is to re-invite a core of 04181 participants, and to reach out more broadly to other members of our respective disciplines, so as to further proselytize the value of cross discipline exchanges so well facilitated by Seminar 04181. We will specifically seek new participants representing other areas which will contribute to a better understanding of Atomicity (e.g. asynchronous hardware design, mobile systems, component technologies). We intend to organize a number of sessions for discussing foundational research, as well as current and future trends in several areas of computer science, such as:
Atomic (ACID) transactions in databases and how they might be applied more widely
Atomic actions in system design and implementation
Atomic broadcast and consensus to increase fault tolerance in many flavors of systems
Atomicity to reason about and define concurrency control
Atomicity in system formalization and analysis
Atomicity in software fault tolerance, error confinement and error recovery
The role of "distributed" Atomicity in systems
Atomicity in application programming, e.g. in workflow systems
Atomicity as an enabling concept that simplifies next-generation systems (AmI, peer-to-peer, etc.)
How Atomicity relates to other concepts used in complex systems

These will be interspersed with cross-area discussion sessions and tutorial sessions aimed at developing a better fundamental understanding of Atomicity and its uses across all our communities. It is our plan to propose case studies and to organize some of the focused discussions around them.

6.2 Preferences: Specification, Inference, Applications

Seminar No. **04271** Date **27.06.–02.07.2004**
Organizers: G. Bosi, R. Brafman, J. Chomicki, W. Kießling

"Preference" is a fundamental notion in those areas of computer science, applied mathematics and philosophy that deal with decisions and choice. In Mathematical Decision Theory, preferences (often expressed as utilities) are used to model people's economic behavior. In Artificial Intelligence, preferences help to capture agents' goals. In Databases, preferences help in reducing the amount of information returned in response to user queries. In Philosophy, preferences are used to reason about values, desires, and duties. Surprisingly, there has been so far very little interaction between those areas. The difference in foci, as well as variations in terminology, make the results obtained in one area difficult to use in another.

This Dagstuhl seminar gathered researchers from many areas involving preferences (in particular databases, AI, mathematics, decision science, philosophy) in order to stimulate more specialized research in those areas and identify possible directions for collaboration. The following topics were covered during the seminar week.

Topics:

- Preference specification and representation
- Preference composition and merging
- Preference aggregation
- Axiomatic properties of preferences
- Logics of preference
- Topological/algebraic preference structures and their utility representation
- Linear and non-linear utility representations
- Preferences with intransitive indifference
- Preference elicitation and learning
- Preference revision
- Incomplete or inconsistent preferences
- Reasoning about preferences
- Priorities in reasoning, conflict resolution and belief revision
- Preference query languages
- Preference query evaluation and optimization
- Preference mining
- Preference repositories
- Preference-driven search engines
- Preference-driven human-computer interaction
- Recommendation systems and other e-commerce applications
- Constraints and preferences, "soft" constraints
- Preference logic programming
- Preference and choice
- Alternatives to preferences

This seminar was unanimously rated as very successful by all participants at the end of a wonderful stay in Dagstuhl. As a main result it was agreed to organize a successor event in the same multi-disciplinary mode. This follow-up event was held as "IJCAI-05 Multi-Disciplinary Workshop on Advances in Preference Handling" in conjunction with the International Conference on Artificial Intelligence IJCAI, July 31 - Aug. 1, 2005, in Edinburgh, Scotland. For 2006 it is planned to continue this tradition as "ECAI-06 Multi-Disciplinary Workshop on Advances in Preference Handling", in conjunction with the European Conference on Artificial Intelligence ECAI in Riva del Garda, Italy, Aug. 28-29.

Chapter 7

Semantics, Specification

7.1 Spatial Representation: Discrete vs. Continuous Computational Models

Seminar No. **04351** Date **22.08.–27.08.2004**
Organizers: R. Kopperman, M.B. Smyth, D. Spreen, J. Webster

Topological notions and methods have been successfully applied in various areas of computer science. Programming language semantics and computing with exact real numbers are two important examples. Computerized geometrical constructions have many applications in engineering. The seminar will concentrate on an important approach which is basic to these applications, i.e. spatial representation.

Due to the digital nature of most applications, the structures used in computer science are different from the mathematical structures that are classically used in engineering and that are based on the continuum. Typical features of these digital structures are asymmetry and partiality. Whereas classical spaces contain only the ideal elements that are the result of a computation (approximation) process, spaces that also allow reasoning on such processes in a formal way must as well contain the partial (and finite) objects appearing during a computation. Only they can be observed in finite time.

The seminar was devoted to the study of several topological structures. The leading example of such is the domain (in Scott's sense), and it is closely related to locales. Here, the finitely observable properties of a process are the primary objects of study. The ideal entities, which are the first class citizens of classical mathematical structures, are obtained as derived objects. These have given rise to a constructive treatment of topological spaces, Formal Topology.

7.2 Semantic Interoperability and Integration

Seminar No. **04391** Date **19.09.–24.09.2004**
Organizers: Y. Kalfoglou, M. Schorlemmer, A. Sheth, S. Staab, M. Uschold

Semantic interoperability and integration is concerned with the use of explicit semantic descriptions to facilitate information and systems integration. Due to the widespread importance of integration, many disparate communities have tackled this problem. They have developed a wide variety of overlapping but complementary technologies and approaches. The seminar has the following objectives:

1. To stimulate collaboration between diverse communities bound by common objectives in the area of semantic interoperability and integration;
2. To lay the foundation for a framework and a theory for understanding and classifying technologies for semantic interoperability and integration;
3. To set the research agenda for this research area the long-term aim of building a "research pipeline" for creating and disseminating results in industry.

Highlights of the Week

Feature talks, short talks, and panel discussions were structured roughly around four main themes: **Mapping and translation, industrial experiences, theoretical foundations, and standards and benchmarks.** Different days were devoted to different themes.

Breakout Sessions

The concluding session on Friday morning was devoted mainly to report back on the discussion and outcomes of the various breakout sessions and in establishing future actions in the field of semantic interoperability and integration.

1. Social Aspects
2. Use Cases and Requirements
3. Mapping Typology and Tools
4. Mapping Notations and Languages
5. Theoretical Foundations
6. Infrastructure and Architectures

Conclusion

We set out to achieve three main objectives in the area of semantic integration and interoperability: 1) to stimulate collaboration, 2) to lay the groundwork for a future comprehensive framework for understanding the field and 3) to set a research agenda.

We successfully brought together experts from industry, academia and government representing historically separate communities including: database integration, category theory,

standards, digital libraries, ontologies, knowledge representation, and the semantic Web. The varied program including invited talks,shorter talks, panels and breakout sessions provided the context for much collaboration. We achieved this objective.

Much of the week was spent wrestling with differences in terminology, identifying a range of relevant broad issues, puzzling over more specific and often subtle distinctions that arise in different sub-areas. The topics for the breakout sessions form a good starting point for the eventual development of a more comprehensive framework that will provide a background for understanding and comparing different techniques, tools and applications that are developed in coming years.

Finally, we considered a range of issues that need to be included on agenda for future research. The problem of semantic interoperability and integration is hard and it is not clear how it could be solved in the near future. For instance, we don't know how to formally specify the problem yet, let alone solving it. Hence it is necessary to work on two fronts: theoreticians need to make their case for the appropriate foundations upon which semantic interoperability and integration can be formalized, while practitioners and users need to expose their local semantics for the benefit of knowledge sharing.

The Semantic Web provides a playground for experimentation, but it also introduces new problems. There are a lot of challenging infrastructure issues still to be addressed and standardization efforts are still at their infancy. Furthermore there is a lack of lengthy experiences and large-scale scenarios to evaluate the scalability of current methods and techniques.

The discussions and outcomes of talks, breakout sessions and panels during the seminar have highlighted these issues and helped to put together efforts which were previously conducted separately in different communities. More discussion and cross-disciplinary collaboration is needed, but the first steps in converging and reaching a consensus might already be well under way.

Chapter 8

Distributed Computation, Networks, VLSI, Architecture

8.1 Peer-to-Peer-Systems and -Applications

Seminar No. **04111** Date **07.03.–10.03.2004**
Organizers: A.D. Joseph, R. Steinmetz, I. Stoica, K. Wehrle

Peer-to-Peer Internet applications have recently been popularized through file sharing applications like Napster, Gnutella, FreeNet etc. Within these applications the Peer-to-Peer networking concept is mainly used to share files, i.e. the exchange of diverse media data, like music, films and programs. The growth in the usage of these applications is enormous and even more rapid than the growth of the World Wide Web.

While much of the attention has been focused on the copyright issues of the shared content, the concept of Peer-to-Peer architectures offers many other interesting and significant research aspects. Due to its main design principle of being completely decentralized and self organizing – opposed to the Internet's "traditional" Client-Server paradigm – the Peer-to-Peer-concept emerges to a major design pattern for future applications, system components and infrastructural services, particularly with regard to scalability and resilience. The perspective of Peer-to-Peer networking offers new challenges, e.g. building scalable and resilient networks and a fast deployment of new services. Based on the decentralized Peer-to-Peer approach new Internet services can be deployed very fast and without spending time-consuming efforts in the process of standardization.

The goal of this Dagstuhl Seminar is to assemble researchers being highly active in the area of Peer-to-Peer mechanisms and networking (1) to reflect on recent research activities, (2) to identify key research issues, i.e. major challenges and (3) to set-up a Peer-to-Peer community in research.

The topics of interest include, but are not limited to:

- Novel Peer-to-Peer applications and systems
- Peer-to-Peer service development

- Peer-to-Peer infrastructure and overlay networks
- Protocols for discovering, management and/or scheduling of resources
- Measurements issues and performance behavior of Peer-to-Peer systems
- Dependability and reliability in P2P networks (fault tolerance, scalability, availability, accessibility, security)
- Anonymity and anti-censorship
- Workload characterization for Peer-to-Peer systems
- Peer-to-Peer mechanisms using other resources than data

The increasing number of research efforts in the area of Peer-to-Peer-networking indicate that there is an enormous interest in and potential for Peer-to-Peer research. This seminar will bring together world-wide leading researchers with next-generation scientists in order to communicate a good picture of the state-of-the-art and to point out directions of future research.

8.2 Wireless Sensor Networks and Applications

Seminar No. **04122** Date **14.03.–19.03.2004**
Organizers: A. Ferscha, S. Olariu, T. Pfeifer

Technical description

The advent of nano-technology has made it technologically feasible and economically viable to develop low-power devices that integrate general-purpose computing with multi-purpose sensing and wireless communications capabilities. It is expected that these small devices, referred to as sensor nodes, will be mass-produced and deployed, making their production cost negligible. Individual sensor nodes have a small, non-renewable power supply and, once deployed, must work unattended. For most applications we envision a massive deployment of sensor nodes, perhaps in the hundreds or even thousands. Aggregating sensor nodes into sophisticated computational and communication infrastructures, called sensor networks, will have a significant impact on a wide array of applications ranging from military, to scientific, to industrial, to health-care, to domestic, establishing ubiquitous wireless sensor networks that will pervade society redefining the way in which we live and work. Sensor networks are currently being established as a specific sub-task of the rapidly unfolding area of ubiquitous and pervasive computing.

The fundamental goal of a sensor network is to produce globally meaningful information from raw local data obtained by individual sensor nodes. Importantly, this goal must be achieved in the context of prolonging as much as possible the useful lifetime of the network and ensuring that the network remains highly available and continues to provide

accurate information in the face of security attacks and hardware failure. The sheer number of sensors nodes in a sensor network, combined with the unique characteristics of their operating environment (anonymity of individual sensors, limited power budget and a possibly hostile environment), pose unique challenges to the design of sensor networks and their applications. For one thing, the limited power budget at the individual sensor node level mandates the design of ultra-lightweight communication protocols. Likewise, issues concerning how the data collected by individual sensor nodes could be queried and accessed and how concurrent sensing tasks could be executed internally are of particular significance. An important guideline in this direction is to perform as much local data processing at the sensor level as possible, avoiding the transmission of raw data through the sensor network. Indeed, it is known that it costs 3J of energy to transmit 1Kb of data a distance of 100 meters. Using the same amount of energy, a general-purpose processor with the modest specification of 100 million instructions/watt performs 300 million instructions. Recent advances in hardware technology are making it plain that the biggest challenge facing the sensor network community is the development of ultra-lightweight communication protocols ranging from training, to self-organization, to network maintenance, to security, to data collection and fusion, to routing, among many others.

There are several possible techniques that can be used to harvest the information produced by a sensor network. Perhaps the simplest involves using one or several sink nodes, special long-range radios, deployed alongside with the sensor nodes. In this scenario, the raw data collected by individual sensor nodes is fused, in stages, and forwarded to the sink nodes that provide the interface to the outside world. Sink nodes are also responsible for training the sensor network and for maintenance and repair operations. However, in some applications, it is impossible or impractical to deploy sink nodes within the sensor network. In such cases the task of harvesting the information produced by the sensor network and that of providing an interface to the outside world may be performed by aircraft and/or helicopters over-flying the sensor network, or by laser transmission to a satellite constellation. In this latter case, the bulk of the inter-sensor communications is by radio, since such communications are point to multi-point, while specialized sensors acting as local sinks communicate with the satellite constellation using laser beams.

While preserved energy can supply short-term applications, sensors dedicated to work over years may need to scavenge energy from the specific environment they are placed into, employing light, temperature, vibration, kinetics, magnetic fields, etc.

The ultra-lightweight protocols may leave not much room for advanced encryption schemes, so protection against overhearing in military applications and privacy protection in personal systems needs to be inherently built into the concepts from the beginning.

Reliability is expected to be a result of the large number of sensors deployed for a specific task. However, this can only be obtained if defective sensors can be excluded from the communication, and the sensors are calibrated – either individually or collectively, either before deployment or continuously in their environment.

Goals of the seminar

The major stated goal of this Dagstuhl seminar is to bring together researchers from these groups and to provide a forum for a stimulating exchange of ideas and cross-fertilization between the participants. We welcome researchers from both industry and academia. Of a special interest to us is the participation of young researchers (35 years or younger). Schloss Dagstuhl is known to offer the ideal setting that will make this experience unique in all respects. We are convinced that the workshop-like relaxed atmosphere, typical of Dagstuhl seminars, will be conducive of high-caliber exchanges of ideas between the participants.

Topics of interest include but are not limited to:

- Energy-efficient training and self-organization
- Failure recovery and recalibration
- Resource management and Connection Admission Control (CAC)
- Media Access (MAC) protocols for wireless sensor network
- Network management scenarios and solutions
- Integration of sensor and terrestrial/satellite networks
- Energy-efficient routing
- Energy and complexity
- Environmental energy scavenging
- Thermodynamics
- Security issues in wireless sensor networks
- Performance evaluation
- Ubiquitous computing
- Applications

The discussions within the seminar are expected to address a large array of problems related to both theoretical and practical aspects of wireless sensor networks and their applications. In particular, presentation of practical implementations, experiments and ongoing projects is strongly encouraged.

8.3 Content Distribution Infrastructure

Seminar No. **04201** Date **11.05.–14.05.2004**
Organizers: C. Griwodz, T. Plagemann, R. Steinmetz

Since the early days of the world wide web (WWW), the information infrastructure provided over the Internet has improved considerably. The simplicity of offering and accessing data on the WWW and the increase of commercial uses of the Internet are major reasons for this development. During this development, it was discovered that shifts in popularity of content and services offered over the web lead to unbalanced load. A sudden increase in popularity can occur (slashdot effect) and is very hard to predict. Therefore it is also very hard to scale these servers and their connection to the Internet at the right time. To overcome these problems, hierarchies of cache servers such as harvest and squid have been set up to alleviated general load, but the large amount of content has limited their effectiveness in many situations. For that reason, controlled pre-distribution of content on behalf of the content providers has been offered as commercial service. To offer the service, servers are installed in the networks of different Internet service providers around the world. The servers can cooperate but they are only connected through the Internet. An infrastructure for this kind of service is known as content distribution network (CDN). Offering content through such a CDN has two advantages. First, several providers' content is offered from the same CDN and only the content of very few providers will experience a sudden increase in popularity at the same time. By offering the content through a CDN, not every content provider must maintain servers that can cope with the load that he experiences at times of very high popularity. Instead, the CDN hosting the content will experience a multiplexed load, and it can be scaled to cope with high load for some but not all content providers. Second, content in the CDN may be replicated over several hosting centers and accesses to content can be redirected to the closest one, thus reducing network load and access latency.

By now, these commercial content hosting CDNs that offer access to discrete media constitute the classical form kind of CDNs. There are, however, other forms of content distribution that are covered by the term CDN as well. One form is used to distribute a different kind of content, namely live data streams that are transmitted from a single source to a large audience. Another form distributes discrete media as well, but uses the peer-to-peer (P2P) model in which equal nodes collaborate for the distribution, and in which nodes are often owned by individuals. Even though commercial CDNs for live and stored content and P2P systems have been investigated separately so far, they share characteristics, for example that they consist of nodes that are connected through the Internet but typically not directly to each other. The integration of these approaches is therefore a challenge to researchers. Another is the increased importance of multimedia content and interactive applications, which will impose new problems on CDNs.

This seminar has brought together researchers who address the challenges that lie in the improvement of CDNs. Among the challenges faced are those related to scaling of the CDN infrastructure, the use of appropriate techniques and tools, and the management of the growing CDNs. Several attendees of the seminar presented ideas and results concerning

this topic. Topics covered in presentations of participants' current research include system and network support for scalability of content and CDN nodes, the communication between nodes and from the nodes to end-systems. The topics included dimensioning, scaling, configuration and reconfiguration of distribution networks for both, hierarchical CDNs and those that follow the P2P model, and reports on the needs of variations applications that rely on CDNs now or in the future.

It became clear that the interaction of end users with services offered via the CDN will become highly important in the future, and that current CDNs can not address this demand in an appropriate manner. Interactivity was therefore identified as a crucial point for the future development of CDNs. Related to this point are danger of denial-of-service attacks on CDNs and questions about the most appropriate support for heterogenous devices and the distribution of applications rather than content.

8.4 Graph Transformations and Process Algebras for Modeling Distributed and Mobile Systems

Seminar No. **04241** Date **06.06.–11.06.2004**
Organizers: B. König, U. Montanari, P. Gardner

Recently there has been a lot of research, combining concepts of process algebra with those of the theory of graph grammars and graph transformation systems. Both can be viewed as general frameworks in which one can specify and reason about concurrent and distributed systems. There are many areas where both theories overlap and this reaches much further than just using graphs to give a graphic representation to processes.

Processes in a communication network can be seen in two different ways: as terms in an algebraic theory, emphasizing their behaviour and their interaction with the environment, and as nodes (or edges) in a graph, emphasizing their topology and their connectedness. Especially topology, mobility and dynamic reconfigurations at runtime can be modelled in a very intuitive way using graph transformation. On the other hand the definition and proof of behavioural equivalences is often easier in the process algebra setting.

Also standard techniques of algebraic semantics for universal constructions, refinement and compositionality can take better advantage of the process algebra representation. An important example where the combined theory is more convenient than both alternatives is for defining the concurrent (noninterleaving), abstract semantics of distributed systems. Here graph transformations lack abstraction and process algebras lack expressiveness.

Another important example is the work on bigraphical reactive systems with the aim of deriving a labelled transitions system from an unlabelled reactive system such that the resulting bisimilarity is a congruence. Here, graphs seem to be a convenient framework, in which this theory can be stated and developed.

So, although it is the central aim of both frameworks to model and reason about concurrent systems, the semantics of processes can have a very different flavour in these theories.

Research in this area aims at combining the advantages of both frameworks and translating concepts of one theory into the other. This workshop was aimed at bringing together researchers of the two communities in order to share their ideas and develop new concepts. The proceedings do not only contain abstracts of the talks given at the workshop, but also summaries of topics of central interest.

8.5 Service Management and Self-Organization in IP-based Networks

Seminar No. **04411** Date **03.10.–06.10.2004**
Organizers: M. Bossardt, G. Carle, D. Hutchison, H. de Meer, B. Plattner

Ad-hoc, peer-to-peer, pervasive, active and programmable networks are emerging research disciplines that pave the way to provision network users with innovative services. However, it turns out that the management of such services is complex, tedious and error-prone. Therefore, applying self-organizing techniques to automate service management and to reduce human intervention is expected to lead to better manageable and more robust network architectures. Hence, the goal of the Dagstuhl Seminar on **Service Management and Self-Organization in IP-based Networks** has been to identify open questions and to discuss new ideas in this exciting research field. The seminar proceedings contain several extended abstracts written by seminar participants, which are briefly introduced in the following sections.

Self-organizing techniques often use **local decision-making** to achieve a **global goal**. The paper by De Meer discusses self-organizing properties of peer-to-peer networks. Network nodes take local decisions based on information received from the network. Papers by Yamamoto and Babaoglu describe mechanisms to distribute information among nodes using gossip- or epidemic dissemination-based approaches. Ripeanu describes an approach to adapt overlay topologies to the state of the underlying network. Papers by Chen, Schulzrinne and Tutschku present network services that adapt to network state (e.g. congestion). Babaoglu presents an interesting example showing how a gossip-based approach can be used to impose a specific network topology. However, it remains an open issue to a priori define the details of the local decision taking process such that it results in a predetermined global outcome.

In nature, many self-organizing systems can be observed and are studied by scientists. Hence, the results of such studies may help engineers to build self-organizing techniques into technical systems. In fact, engineers have developed bio-inspired algorithms that mimic the behavior of biological systems to solve a variety of problems. Dressler intends to secure networks using an architecture directly derived from cell and molecular biology. Tschudin proposes a method to implement self-healing communication protocols. While Dressler suggests a technical system mimicking nature as closely as possible, Tschudin's approach is rather inspired by the resulting behavior of biological systems. Considering these approaches that mark two opposite points in the design space of bio-inspired systems,

it is not yet fully understood to what extent technical approaches should mimic biological systems in order to achieve optimal results.

The different flavors of self-organizing techniques presented above aim at simplifying the management of networks. The question arises, which are the primitives, knobs and dials that network architectures should offer to support their implementation. Sterbenz discusses where knobs and dials should be placed. Papers by Calvert, Hjlmtýsson, Wehrle and Sifalakis describe network architectures providing new primitives that simplify the implementation of new functionality within the network. Approaches described by Brunner, Farkas, and Minden aim at providing tools and mechanisms to dynamically deploy new functionality in networks, while the paper by Ruf presents a flexible platform to efficiently execute such functionality. Finally, papers by Gao and Gerke describe economical aspects of service management in new network architectures.

The Dagstuhl Seminar brought together 38 researchers and engineers from Australia, USA and Europe. Their different background and numerous presentations resulted in interesting discussions. The organizers are very grateful to all the participants for actively participating, raising questions and providing new insights. We would like to thank James Sterbenz for supporting us during the conference and the Dagstuhl Association for providing our research community with an excellent venue for scientific exchange.

8.6 Cognitive Networks and Radios

Seminar No. **04431** Date **17.10.–22.10.2004**
Organizers: J. Mitola, P. Mähönen, J. Pereira

Software and Cognitive Radios

In the communications engineering the software radios have become an established paradigm during the last decade or so [Mitola 1995; 1999]. There is a large body of research especially in Europe and in the USA towards realizing more complex and efficient software radio prototypes. The original term of software radio was introduced by J. Mitola in early 1990's. The basic idea behind the software radio is that increasing computing power and enhanced software algorithms especially for digital signal processing are leading us to a situation in which most of the wireless communication functionalities will be handled by powerful processors running software instead of highly customized hardware. The ultimate software radio would include only smart antennae, AD/DA-circuits, and everything else would be done through software in powerful processing unit. It has pointed out that "As the software radio makes its transition from research to practice, it becomes increasingly important to establish provable properties of the software radio architecture on which product developers and service providers can base technology insertion decisions. Establishing provable properties requires a mathematical perspective on the software radio architecture" [Mitola 1999]. The work on software radio is also conducted quite often under the theme "reconfigurable radio" or "reconfigurable communication technologies".

It has been, of course, generalized to consider not only the wireless radios, but also networking paradigms. In some sense, active networks or programmable networks share a same idea space with software radios.

A far more recent paradigm is so-called cognitive radio introduced by Mitola and Maguire. The basic idea of cognitive radio is deeply embedded with the context sensitivity and context sensitive inference. The context sensitivity in the case of ubiquitous computing and pervasive computing has been a long time known within the computer science community (just to mention the seminal suggestions by Bush, Engelbert and Weiser are enough). The current challenge with cognitive radios is not only technology itself. In the case of cognitive radios one of the challenges is to foster interdisciplinary research, and to understand more clearly what are the specific research problems and boundary conditions that should be met. As an example, it is clear that machine learning and artificial intelligence methods have their important part to play with cognitive radios. Cognitive radio technology is also quite often much harder to achieve than in the case of typical context sensitive applications; stringent real-time requirements and need to co-operate between different protocol stacks make it a complex task, even to model.

Cognitive Networks

As the cognitive radio is an extension of software radio (and active networking could be seen as networking counter part of programmable radio), we will speculate with a term cognitive networks . The cognitive network as an idea has been introduced, e.g. by D. Clark [2003] from MIT and Craig Partridge [2003], both well known initial major contributors towards the development of Internet. Clark has described knowledge plane paradigm for Cognitive Networking. Some European researchers have also tried to map challenges and possibilities that could be encountered with wireless and mobile cognitive networks. There has been also some activity emerging from machine learning community in the U.S.A. The cognitive networks have a real chance to become a new architectural paradigm change. The current activity, and admittedly scarce, first initial results show that cognitive network might become a very strong research line in the future. It shares in part same challenges as cognitive radios also in the research organization sense. Definitely a lot of initial work is required to nail down the clear research issues and challenges in order to avoid unnecessary "hyped" work.

Goals of the Seminar, Participants, Industry

All of these focus areas are emerging fields (software radios are established, but not really in markets, cognitive radios are rising field, and in the case of cognitive networks we really do not know clearly how to define the boundaries of the research field). Because of these, the seminar could be seen as a slightly unconventional research retreat and workshop. It is not organized "just" to describe and talk about recent results. It is more like (beginning of) exploration – or meeting of explorers to agree where we should go. We aim at in the seminar to

- Define more clearly what should be meant with cognitive radios and networks research.
- To understand what are the research challenges, and what should be done to meet them. (Output should be, e.g., a list on "What are the problems we would like to present to experts in different research fields?" We need to communicate many challenges to other fields from the pure communications sciences and engineering domain. But first, we have to make clear what those challenges are.)
- More specifically; what are the scalability and complexity boundaries for such systems (in the system engineering and network engineering sense), including the realistic possibility for embedded solutions. Many solutions exist already, but in practice they do not scale for communications systems. The more specific questions we need to consider are
  - scalability
  - architecture and topology constraints
  - implementability
  - specific research problems esp. cross-layer protocol optimization etc.
  - machine learning, optimization and AI related issues
  - roadmap for progress, and hardware boundaries.

If we are able to provide some answers and definitions, we feel that we have succeeded. Due to this we are reserving a beginning of the time for a small number of speakers (including a keynote from Chuck Thacker). All the invited members are also welcomed to suggest a contributing paper or short talk to this Dagstuhl seminar.

The main format for our 3,5 day meeting is to have productive workshop and brainstorming. The idea is that collectively we could produce a sort of "consensus report" on the future of cognitive networking, and roadmap towards it. We will have a small number of bright graduate students to make notes during the brainstorming sessions, if so required. Also as a sub-goal would be to discuss that should we have joint research projects to do actual work? The very nice environment and informal atmosphere of Dagstuhl should help on this, and we have reserved plenty of time for informal discussions. A good number of people from industry are also participating in. The informal networking between different academics and some industry people is planned to foster possible future research projects. Results of our brainstorming and all contributed short papers and talks will be collected to seminar report (proceedings) that will be distributed to all participants.

8.7 Mobile Information Management

Seminar No. **04441** Date **24.10.–29.10.2004**
Organizers: M. Dunham, B. König-Ries, E. Pitoura, P. Reiher, C. Türker

From October, 25th until 29th, 2004, a Dagstuhl seminar on mobile information management took place. The seminar was attended by 34 researchers from Europe and North America with backgrounds ranging from database systems, mobile information systems, geographic information systems and business informatics to wireless networks and security.

The motivation for organizing this seminar was as follows: The combination of wireless and wired connectivity along with increasingly small and powerful mobile devices, such as laptops, personal digital assistants, handheld/tablet PCs, and smart phones, but also embedded devices and sensors, enables a wide range of new applications and new ways to use existing applications. This will radically change the way information is managed and processed today. Information becomes ubiquitous, highly distributed and at the same time accessible from everywhere at any time. Information access takes place in highly dynamic and instable networks. Nevertheless, users and application developers will expect information processing to continue under similar guarantees as those offered by todays stationary and more or less centralized systems, even if some nodes of the ubiquitous information network are (temporarily) disconnected and/or are in motion. Examples for these guarantees are the ones given by database management systems, i.e. consistency and durability of data. Additional challenges stem from the fact that mobile devices while evolving fast, are (and will always be) less powerful than their stationary counterparts; they are smaller (resulting, e.g., in smaller in- and output devices and less available storage), have restricted energy supplies and communicate via expensive, unreliable wireless communication media and require new solutions to major existing security problems.

The goal of the seminar was to bring together researchers, especially from the area of databases and information systems, to identify open problems and new challenges in data, service, and user management in mobile information processing environments. We were particularly interested to provide a forum for discussing the consequences of the mobility of users and devices on today's and future data management systems. The first aim of these discussions was to clearly identify these consequences. The second and more challenging aim was to determine where existing solutions can be applied, where mobility raises truly new challenges, and which of these challenges are there to last. Of course, these discussions were to take into consideration not only general solutions from the database community at large, but also the approaches that have been developed by the mobile databases community over the last decade, e.g.:

- Moving objects and mobile users
- Mobile data dissemination and delivery
- Mobile data replication and synchronization
- Discovery and composition of mobile services
- Mobility awareness and adaptability
- Location-dependent, context-based querying and optimization
- Designing location-aware, context-aware services

- Mobile sensor and stream data management
- Continuous querying
- Self-organizing, self-tuning mobile components
- Ad-hoc processes and networks
- Mobile transactional processing
- Quality of Service for mobile databases
- Low power cryptography and other cryptographic solutions designed for wireless networks
- Methods of identifying, authenticating, and safely integrating mobile devices into a network

The discussions in this seminar were supposed to result in the identification of promising avenues of research that should be pursued in order to address the new issues that are introduced by mobility.

In order to achieve the goals mentioned above and in particular to foster discussions, the seminar was organized along two main lines: First, a number of overview talks were given. Second, a considerable amount of time was spent working in smaller groups of five to fifteen participants. Here, a number of topics were looked into more detail, research challenges and open issues were identified, avenues for future research were looked for.

Working Groups

During the seminar, most of the work was done in working groups. We had two sets of three working groups each. Each group got together for about a day, working on a more (or in one case: less) specific topic. Working group sizes ranged from five to fifteen participants. Working group sessions were concluded by a presentation of the results to the plenum. The following working groups met.

WG1: Mobile Business. This working group tried to get a better understanding of the key players in mobile business, how powerful they are and what their relationships are. A clear understanding of this system is necessary to enable successful development and deployment of mobile applications.

WG2: Mobile Transactions. The group looked at different scenarios in which mobile transactions are needed, ranging from infrastructure-based networks to ad-hoc networks. Despite the large amount of work that has been done on mobile transactions in the past, a number of challenging research questions remain to be addressed, in particular with respect to networks with little or no infrastructure.

WG3: Mobile Queries. After classifying different kinds of mobile queries and identifying the different dimensions that make mobile query processing so challenging, the working

group concentrated on the role of context to ease dealing with mobile queries. There, a number of open research questions were highlighted.

WG4: Mobile Application Design. This working group started with the question "Why is the development of mobile applications harder than that of "normal" software?" In an attempt to answer this question, a number of mobile application scenarios and common experiences with the development were collected. It became evident that mobile application design requires to take a high number of dimensions, most of which are somewhat interrelated into account simultaneously. The group drew up an initial list of these dimensions.

WG5: MANETs, P2P and Self-Organization. This working group addressed a rather broad set of questions. After identifying similarities and differences between the three classes of systems regarded, the group concentrated on identifying open research issues in all three areas.

WG6: Description and Matching of Services in Mobile Environments. The working group classified services in mobile environments along two dimensions: The mobility or non-mobility of the service provider and the location-dependence or independence of the service offered. They suggest that this classification should be reflected in the service description. Therefore, they developed an initial idea how service descriptions should be structured to be suitable in mobile environments: Descriptions should be split into a static part (containing the regular service description) and a dynamic part (describing the current context of the service provider). Depending on the class of service, the dynamic part will be more or less complex and important. Analogously, service requests should be split into several parts. Again, depending on the class of offer and request, different algorithms for matching of offers and requests need to be used. While there was general agreement that the idea seemed plausible, its realization will require considerable research effort.

Results of the Seminar

Mobile information management is a topic that is of immense and growing interest to a number of communities. It seems important to bring these communities together to ensure the development of appropriate solutions. While the databases community certainly is strong in questions of mobile information management and has a lot of solutions to offer, input from other researchers is needed: For instance, security issues are neglected in a frightening way in today's mobile applications and in particular in research prototypes. If we want these developments to enter the mass market, this needs to change. Research is often carried out without knowledge about or focus on existing business models etc. This, too, obviously, hampers market success.

A second result of the seminar was the identification of a number of promising avenues of research. These were discussed in more detail in the working groups.

Mobile information management is clearly an area of growing importance, and the research community should focus greater attention on providing the kinds of systems and services required to allow the field to reach its potential for creating new industries and improving

people's lives. Mobility is not just an add on, but brings with it challenges that are different enough from more traditional system to warrant dedicated research.

Also, research on mobile data management offers a solid foundation for work on peer-to-peer and self-organizing systems and ultimately for ubiquitous computing. Again, this indicates the importance of this research area.

8.8 Future Generation Grids – FGG 2004

Seminar No. **04451** Date **01.11.–05.11.2004**
Organizers: M. Cosnard, V. Getov, D. Laforenza, A. Reinefeld

The Internet and the Web have had a major impact on society. By allowing us to discover and access information on a global scale, they have created entirely new businesses and brought new meaning to the term "surf". Yet, simply being able to offer and access information on the Web is ultimately unsatisfactory: We want processing and, increasingly, we want collaborative processing within distributed teams. This need has led to the creation of the Grid – an infrastructure that enables us to share capabilities, integrate services and resources within and across enterprises, and allows active collaborations across distributed, multi-organizational environments. Powered by on-demand access to computer resources, seamless access to data, and dynamic composition of distributed services, the Grid promises to enable fundamentally new ways of interacting with our information technology infrastructure, of doing business, and practicing science. It represents perhaps the final step in the great disappearing act that will take computing out of our homes and machine rooms into the fabric of the society, where it will stand alongside telephone switches, power generators, and the other invisible technologies that drive the modern world. Future applications will not only use individual computer systems, but a large set of networked resources. This scenario of computational and data grids is attracting a lot of attention from application scientists as well as from computer scientists. In addition to the inherent complexity of current high-end systems, the sharing of resources and the transparency of the actual available resources introduce not only new research challenges but also a completely new vision and approaches to designing, building, and using future generation Grid systems.

Overview of the Seminar

The seminar brought together 45 scientists and researchers in the Grid area in an attempt to draw a clearer picture of future generation Grids and to identify the most challenging problems on the way to achieving the invisible information Grid ideas in our society. The participants came from France (12), Germany (10), Italy (8), Great Britain (5), The Netherlands (3), Belgium (1), Cyprus (1), Czech Republic (1), Poland (1), Spain (1), Switzerland (1), and the U.S.A. (1). This was the first seminar in a series of workshops planned by the just started EU Network of Excellence project CoreGRID the "European Research Network on Foundations, Software Infrastructures and Applications for large

scale distributed, GRID and Peer-to-Peer Technologies". The CoreGRID Network of Excellence aims at strengthening and advancing scientific and technological excellence in the area of Grid and Peer-to-Peer technologies. To achieve this objective, the Network brings together a critical mass of well-established researchers (119 permanent researchers and 165 PhD students) from forty-two institutions, which have constructed an ambitious joint programme of activities. Additional impetus for the organization of the FGG-Seminar came from another EU project, the "ERA Pilot on a Coordinated Europe-Wide Initiative in Grid Research" (GridCoord). It targets at strengthening Europe's position on grid research and its exploitation by overcoming the fragmentation and dispersion across the EU research programs. The FGG seminar helped in getting an overview on the various Grid initiatives and projects and thereby provided a good basis for drafting a compendium on Grid research programmes in major European states. Background information for the seminar came from the just recently published findings of a EU expert group on Next Generation Grids. (http://www.cordis.lu/ist/grids/index.htm)

Results

The talks of the seminar were grouped into the broad topics middleware, software toolkits, Grid and peer-to-peer system architecture, data and information management, resource management, and scheduling. In an attempt to provide an overview on the status of the various national Grid Initiatives – a topic deemed important especially for the GridCoord project – the following Grid Initiatives were presented:
Grid.it (Italy)
D-Grid (Germany)
DAS-2 (The Netherlands)
SGIGrid (Poland)
UK e-Science (UK)
ACI GRID's Grid 5000 project (France)

While the general goal of establishing a national Grid for the benefit of science and research in the respective countries is similar, each of these initiatives puts an emphasis of slightly different aspects. Most apparent is perhaps the "virtual laboratories" approach in the Netherlands, the more experimental character of the French Grid 5000 project as part of the ACI GRID initiative and the strong trend towards the deployment of productive application scenarios in the UK e-Science initiative. However, it is difficult to summarize the subtle differences in the initiatives in this brief summary and therefore, a more detailed analysis must be left for the future. The discussion session on next generation Grid technologies focussed largely on the importance of making Grid systems "autonomic" in the sense that future Grid components shall be able to autonomously cope with failures without affecting the other "healthy" components. Even more emphasis was put on the discussion of the newly established Web Services Resources Framework (WSRF) versus the previous Open Grid Service Infrastructure (OGSI), Web Services, and Service Oriented Architectures (SOA) in general.

Chapter 9

Modelling, Simulation, Scheduling

9.1 Component-Based Modeling and Simulation

Seminar No. **04041** Date **18.01.–23.01.2004**
Organizers: F.J. Barros, A. Lehmann, P. Liggesmeier, A. Verbraeck, B.P. Zeigler

With respect with permanent decreasing innovation cycles of increasingly complex distributed systems, and regarding their efficient operation and maintenance, application of quantitative models and simulations is of growing importance

- for systems design and development,
- for configuration planning, and operation and maintenance management and
- for training and education.

Effective model and simulation applications require adaptability and reusability of models and submodels. Therefore, conceptual and methodological approaches for modular, component-based model construction (as opposed to complex monolithic models), and adaptation of model components – based on solid theoretical foundations – is urgently required. The lack of system theoretic foundations for component-based modeling and simulation, and missing international standards, as well, currently results in the implementation of proprietary models, model components and component interfaces, allowing only limited interoperability and reusability of those modules.

Regarding the above mentioned trends and requirements major topics of this Dagstuhl-Seminar will focus on:

- System theoretic definitions and foundations for model-components (especially regarding differences to definitions and specifications of software-components),
- Specification of model components (and interfaces),
- Hierarchical, component-based model development regarding cost-benefit, quality, performance, reliability, and reusability aspects,

- Architectures for component-based model and simulation applications,
- Verification and validation methods for model components,
- Synergies for the application of software-component-technologies in their application for building model-components.

9.2 Scheduling in Computer and Manufacturing Systems

Seminar No. **04231** Date **31.05.–04.06.2004**
Organizers: J. Błażewicz, K. Ecker, E. Pesch, D. Trystram

The biannual Workshop on "Scheduling in Computer and Manufacturing Systems" was organized in Dagstuhl Castle during the week May 31-June 4, 2004. It was the fifth meeting and gathered 58 participants from universities or research centers in Austria, Belarus, Belgium, Brazil, Canada, France, Germany, Great Britain, Hong Kong, Hungary, Israel, Italy, Poland, Switzerland, The Netherlands, and the U.S.A.

The objective of the seminar was to provide a forum for the discussion of ongoing research in scheduling. The seminar promoted an exchange of ideas covering the entire spectrum from case studies of real applications to recent advances in mathematical foundations. The various aspects of scheduling were covered by 39 lectures that addressed classical application areas such as distributed processing, operating systems, dependable systems, and flexible manufacturing. It is worth pointing out that many lectures were motivated by practical considerations, as for example machine break-downs, batch scheduling, synchronous production, robotic cell scheduling, real-time scheduling, and resource investment problems. But also exciting new areas emerged such as those in modern communications systems, examples being wireless networks, multimedia networks, and the internet.

The seminar proceeded along three broad fronts: applications, which include empirical studies of existing systems as well as numerical studies of the analysis and simulation of system models. Most of the application studies came from the area of production scheduling and planning, such as just-in-time scheduling, due date assignment and project control, including special problems dealing with machine break-downs, robotic cells, assembly scheduling, load balancing, minimizing the number of workers (human resources). Other presentations considered special problems from chemistry and oceanography, the design of schedulers, e.g. for web applications, and planning examination sessions. Algorithms were presented for various problems such as batch scheduling, resource scheduling, tardiness problems, shop problems, deadline and due date scheduling, real-time scheduling, on-line scheduling, single machine problems, time lags, scheduling with communication delays. The main concern in these presentations was the design and analysis of algorithms ranging from simple and tractable on-line and greedy rules to methods based on semi-enumerative approaches, branch and bound, local neighborhood search, and LP formulations. New theoretical developments included recent results in the analysis of new and classical problems

under novel (or multiple) criteria, dealing with particular assumptions on machines, tasks (e.g., release dates, precedence constraints, communication delays, multiprocessor tasks, bi-processor tasks), and other problems such as assembly scheduling problems and on-line scheduling. Typical questions discussed were the structure of problems and their relation to graph theory, complexity of problems including polynomial solvability, the design of algorithms and performance analysis, and the approximability of optimal solutions.

Chapter 10

Data Bases, Information Retrieval

10.1 Content-Based Retrieval

Seminar No. **04021** Date **04.01.–09.01.2004**
Organizers: J. Malik, H. Samet, R. Veltkamp, A. Zisserman

Images, music, video, and 3D scenes play a crucial role in Visual Information Systems and Multimedia. There is an extraordinary number of applications of such systems in entertainment, business, art, engineering, and science. Such applications often involve huge collections of media, so that efficient and effective searching in databases of these media is an important operation.

This seminar is the follow-up of the Dagstuhl Seminars on Content-Based Image and Video Retrieval in December 1999 and January 2002. One of the notable trends was the extension of the retrieval process from images and video, to also 3D scenes and music, often integrated with text. Many of the same algorithmic elements can be seen in the retrieval of these different media. We have therefore decided to broaden the scope of the seminar, and call the seminar "Content-Based Retrieval".

The emphasis of this third seminar will lie on the algorithmic aspects of all kinds of content-based retrieval. Fundamental questions such as feature extraction, pattern similarity rating, indexing large collections, and the role of perception play an important role. We strongly believe that content based retrieval needs an integrated approach from fields such as image processing, shape processing, perception, data base indexing, visualization, querying, etc. The purpose of this seminar is to bring together people from the various fields in order to promote information exchange and interaction among researchers who are interested in various aspects of accessing the content of images, music, video, and 3D data.

10.2 Detecting Local Patterns

Seminar No. **04161** Date **12.04.–16.04.2004**
Organizers: J.-F. Boulicaut, K. Morik, A. Siebes

Introduction

The dramatic increase in available computer storage capacity over the last 10 years has led to the creation of very large databases of scientific and commercial information. The need to analyse these masses of data has led to the evolution of the new field *Knowledge Discovery in Databases (KDD)* at the intersection of machine learning, statistics and database technology (Fayyad et al, 1996). Being interdisciplinary by nature, the field offers the opportunity to combine the expertise of different fields to a common objective. Moreover, within each field diverse methods have been developed and justified with respect to different quality criteria. It has to be investigated, in which way these methods can contribute to solving the problem of KDD.

Traditionally, KDD was seeking to find global models for the data that explain most of the instances of the database and describe the general structure of the data. Examples are statistical time series models, cluster models, logic programs with high coverage or classification models like decision trees or linear decision functions. In practice, though, the use of these models often is very limited, because global models tend to find only the obvious patterns in the data, that domain experts already are aware of (Guyon et al., 1996). What is really of interest to the users are the local patterns that deviate from the already known background knowledge. The new field of *local patterns* has been proposed by David Hand who organised a workshop in 2002.

The Dagstuhl Seminar on Local Pattern Detection brought together experts from Europe, Japan, and the United States – 13 countries were represented. Moreover, the participants brought with them expertise in the following fields: Decision trees, Regression methods, Bayesian models, Kernel methods, Inductive Logic Programming, Deductive Databases, Constraint Propagation, Time Series Analysis, Query Optimisation, Outlier Detection, Frequent Set Mining, and Subgroup Detection. All talks were focused on the topic of local patterns in order to come to a clearer view of this new field.

Novelty of Local Pattern Detection

Research has investigated global models for a long time in statistics and machine learning. The database community has inspected the storage and retrieval of very large datasets. When statistical methods encounter the extremely large amount of records and the high dimensionality of the stored observations, exploratory methods failed. Machine learning already scaled up to build-up global models, either in the form of complete decision functions or in the form of learning all valid rules from databases. However, the classification does not deliver new, surprising insights in the data, and the valid rules reflect almost exactly the domain knowledge of the database designers. In contrast, what users expect from the exploratory analysis of databases are new insights into their data. Hence, the matter of interestingness has become a key issue.

The success of Apriori or subsequently frequent set mining can be explained by it being the first step into the direction of local patterns. The correlation of more than the few features, which standard statistics could analyse, could successfully be determined by frequent set

mining. Frequent set mining already outputs local patterns. Current research tasks within this set of methods are algorithmic concerns as well as the issues of interestingness measures and redundancy prevention. The collaboration of database specialists and data miners has led to the notion of inductive databases. The new approach writes measures of interest and the prevention of redundancy in terms of constraints. Also users can formulate their interests in terms of constraints. The constraints are pushed into the search process. This new approach was discussed at the seminar intensively and it was found a view covering diverse aspects of local patterns, namely their internal structure and the subjective part of interestingness as given by users. Talks on frequent set mining were presented by

- Francesco Bonchi and Fosca Giannotti showing the use of constraints within the search for local patterns,
- Rosa Meo presenting a language for inductive queries expressing constraints,
- Jean-Francois Boulicaut applying frequent set mining to gene expression data by exploiting Galois operators and mining bi-sets, which link situations and genes,
- Cline Rouveirol reporting on the combination of frequent sets found in gene expression and genome alteration data,
- Bart Goethals offering a new constraint on the patterns, namely that of s% of the database containing the minimal number of tiles, where each tile has the maximal number of 1s..

Subgroup discovery has had already some successes. Stefan Wrobel and Nada Lavrac reported on theory and applications of subgroup discovery, in particular focusing on interest measures and the significance of patterns found.

- Stefan Wrobel clearly indicated the problem of false discoveries and presented two approaches: the MIDOS algorithm, which finds subgroups according to the true deviation, and a sequential sampling algorithm, GSS, which makes subgroup discovery fast. He also tackled the redundancy problem by maximum entropy suppression effectively. Applications on spatial subgroup discovery concluded the talk.
- Nada Lavrac reported on successful applications of subgroup mining in medicine.
- Josef Frnkranz presented a unifying view of diverse evaluation measures.

The statistical view was presented in five talks:

- Xiaohui Liu builds a noise model using supervised machine learning methods and after that local patterns are detected. Testing them against the noise model yields clean data. The approach was illustrated with two biomedical applications.
- Niall Adams and David Hand distinguish two stages in pattern discovery

    1. Identify potential patterns (given a suitable definition)

2. Of these, identify significant (in some sense) patterns (expert or automatic)

They noticed that the former is primarily algorithmic and the latter has the potential to be statistical. They illustrated this with an application on discovering cheating students.

- Claus Weihs focused on the transformation of local patterns into global models illustrated with the transcription of vocal time series into sheet music
- Frank Hppner discussed the similarities and differences between clustering and pattern discovery. In particular he showed how interesting patterns can be found by the clever usage of a hierarchical clustering algorithm.
- Stefan Rping introduced a general framework in which local patterns being produced by different processes are identified using a hidden variable. This allows for the use of the EM algorithm to discover the local patterns directly, that is, without reference to the global data distribution. A new scaling algorithm handles the combination of classifiers. The method was illustrated with business cycle data.

Local patterns need to have an internal structure in order to be meaningful, interesting, and distinguished from noise. Several talks focused on internal structures:

- Arno Siebes employed a graphical view on data and patterns to express this internal structure. Moreover, aggregate functions along paths in these graphs were used to compute new features.
- Katharina Morik discussed the importance of the example representation LE, because it determines the applicability of methods. For local pattern detection frequency features are best suited. She showed how to characterize time-stamped data using a frequency model.
- Helena Ahonen-Myka gave an overview of sequence discovery with a focus on applications on text.
- Myra Spiliopoulou gave an overview of local patterns exhibiting temporal structures.
- Thorsten Joachims investigated internal structures such as parse-trees and co-reference pairing. He presented a general method how such structures can be analyzed by SVMs. Moreover he showed how the combinatorial explosion of the number of constraints can be controlled by the upper bounds derived from statistical learning theory.

Seminar Results

Based on the definition of David Hand (2002)
data = background model + local patterns + random
seminar participants came up with 12 definitions, what local patterns actually are. These were intensively discussed and we finally agreed on the following.

- Local Patterns cover small parts of the data space. If the learning result is considered a function, then global models are a complete function, whereas local patterns are partial.

- Local Patterns deviate from the distribution of the population of which they are part. This can be done iteratively – a local pattern can be considered the overall population and deviating parts of it are then determined.

- Local Patterns show some internal structure. For example, correlations of features, temporal or spatial ordering attributes, and sequences tie together instances of a local pattern.

Local patterns pose very difficult mining tasks:

- Interestingness measures (differ from standard criteria for global models)

- Deviation from background knowledge (global model) asks for good estimates of the global mode, where local patterns deviate from the overall distribution

- Modeling noise (for data cleaning, distinguish from local patterns)

- Automatic feature generation and selection for local patterns (for local patterns other features are successful than for global models, standard feature selection does not work)

- Internal structures of the patterns (correlation of several features, graph, sequence, spatial closeness, shape) can be expressed in several ways, e.g., TCat, constraints.

- Test theory for an extremely large space of possible hypotheses (large sets are less likely, hence global models do not encounter this problem)

- Curse of exponentiality - complexity

- Redundancy of learned patterns

- Sampling for local patterns speeds up mining and enhances quality of patterns

- Evaluation: benchmark missing

- Algorithm issues

The speakers at the workshop have been invited to submit a chapter for edited post-workshop proceedings.

10.3 Perspectives Workshop: Data Mining: The Next Generation

Seminar No. **04292** Date **11.07.–16.07.2004**
Organizers: R. Agrawal, J.C. Freytag, R. Ramakrishnan

Data Mining has enjoyed great popularity in recent years, with advances in both research and commercialization. The first generation of data mining research and development has yielded several commercially available systems, both stand-alone and integrated with database systems; produced scalable versions of algorithms for many classical data mining problems; and introduced novel pattern discovery problems.

In recent years, research has tended to be fragmented into several distinct pockets without a comprehensive framework. Researchers have continued to work largely within the parameters of their parent disciplines, building upon existing and distinct research methodologies. Even when they address a common problem (for example, how to cluster a dataset) they apply different techniques, different perspectives on what the important issues are, and different evaluation criteria. While different approaches can be complementary, and such a diversity is ultimately a strength of the field, better communication across disciplines is required if Data Mining is to forge a distinct identity with a core set of principles, perspectives, and challenges that differentiate it from each of the parent disciplines. Further, while the amount and complexity of data continues to grow rapidly, and the task of distilling useful insight continues to be central, serious concerns have emerged about social implications of data mining. Addressing these concerns will require advances in our theoretical understanding of the principles that underlie Data Mining algorithms, as well as an integrated approach to security and privacy in all phases of data management and analysis.

Researchers from a variety of backgrounds assembled at Dagstuhl to re-assess the current directions of the field, to identify critical problems that require attention, and to discuss ways to increase the flow of ideas across the different disciplines that Data Mining has brought together. The workshop did not seek to draw up an agenda for the field of data mining. Rather, it offers the participants' perspective on two technical directions – compositionality and privacy – and describes some important application challenges that drove the discussion. Both of these directions illustrate the opportunities for cross-disciplinary research, and there was broad agreement that they represent important and timely areas for further work; of course, the choice of these directions as topics for discussion also reflects the personal interests and biases of the workshop participants.

10.4 Perspectives Workshop: Data Warehousing at the Crossroads

Seminar No. **04321** Date **01.08.–06.08.2004**
Organizers: J. Hammer, M. Schneider, T. Sellis

Motivation

Research in data warehousing and online analytical processing (OLAP) has produced important technologies for the design, management and use of information systems for decision support. Much of the interest and success in this area can be attributed to the need for software and tools to improve data management and analysis given the large amounts of information that are being accumulated in corporate as well as scientific databases. However, despite the continued success and maturing of the field, much work remains to be done across many different areas of data warehousing. As more and more information is managed and stored electronically, data warehouses continue to increase in size at a staggering rate. At the same time, the data that needs to be stored in the warehouse is getting more complex in both structure and semantics while the analysis must keep up with the demands of new applications. For example, customer relationship management, analysis of spatial and spatio-temporal data, OLAP mining, mobile OLAP and more recently applications in life sciences are demanding novel representation and manipulation techniques for non-standard data, efficient algorithms to compute aggregate queries, and new, application-specific index structures.

Goal

The goal of this seminar is to bring together researchers, especially from the area of databases and information systems, to review the state-of-the-art in data warehouse technologies for OLAP, discuss recent advances and trends in the field, and identify interesting research problems. An important final outcome will be a research agenda describing opportunities and challenges for what the participants deem to be the most promising new areas in data warehousing research. The write-ups describing the topics will be jointly authored by the organizers and seminar participants and published in the form of an edited book for dissemination to the broader community.

Organization and Contents

The seminar will consist of a mixture of short presentations to the entire group as well as parallel working sessions where specific technical issues will be discussed in greater depth. In order to help the organizers in the planning of the sessions, all participants are asked to prepare and submit a one page position statement identifying one or more research issues that should be addressed in the seminar. In order to achieve maximal success, position papers should focus only on warehousing technologies supporting non-standard or emerging applications such as analysis of spatial, spatio-temporal, Web, semistructured, (bio-)medical, and bioinformatics data just to name a few. Topics include but are not limited to:

- Conceptual modelling for data warehouses (multidimensional data modelling)
- Design methods (logical design, physical organization, tuning)

- New data warehouse architectures such as XML warehouses, Web warehouses, or data warehouse federations
- Integration of warehouses into the business processes of an enterprise
- Management of data quality as well as uncertainty, vagueness, and imprecision
- Organization of meta-data management
- Multidimensional query languages
- Query processing and optimization
- Storage and indexing methods
- Data warehouse operational processes, including novel Extract-Transform-Load (ETL) tools
- Data warehouse evolution

The position papers are due on 01/07/2004 and will be used by the organizers to identify a preliminary list of research topics as well as a temporary assignment of participants into working groups. Topics and group assignments will be finalized on the first day of the seminar. We expect there to be approximately 8-10 different groups which will be charged to work out the research agenda for their assigned topic. Morning sessions will be used for status reports by the individual groups and to further cross-fertilization among related topic areas. Group working sessions will be scheduled primarily during the afternoons. To help prepare for the technical discussions, all participants will receive a copy of everybody's position statement as well as the temporary group assignment prior to their arrival in Dagstuhl.

Chapter 11

Other Work

11.1 Theory of Evolutionary Algorithms

Seminar No. **04081** Date **15.02.–20.02.2004**
Organizers: H.-G. Beyer, T. Jansen, C. Reeves, M.D. Vose

The 2004 Dagstuhl seminar on the theory of evolutionary algorithm was the third following seminars in 2000 and 2002 which had the same title. These Dagstuhl seminars are recognized within the evolutionary algorithm community as a biannual series of high quality meetings concerned with theoretical aspects of evolutionary computing. Therefore, there was great interest in participating and presenting ideas. The seminar had 49 participants from twelve different countries, namely eighteen from Germany, eleven from the United States, eight from Great Britain, three from the Netherlands, two from Canada and one each from Austria, Australia, France, Japan, Mexico, Russia, and Switzerland.

From these 49 researchers, 39 were able to give a presentation. The topics of the talks were as diverse as the current status of evolutionary computation theory. There have been talks on general aspects of modeling and analyzing evolutionary algorithms, in particular talks on the pros and cons of infinite population approaches. Also different ways of classifying and analyzing landscapes defined by different problems and genetic operators have been presented. Many talks presented concrete analytical results on specific evolutionary algorithms and specific problems. Such results have been presented for both, continuous and discrete search spaces, for artificial problems or problem classes as well as for a typical combinatorial optimization problem. Recent trends like optimization of noisy objective functions, estimation of distribution algorithms, and coevolution have been topic of several talks as well as the well-established genetic programming. Other randomized search heuristics used for optimization have also been subject of talks, namely estimation of distribution algorithms and ant colony optimization.

The talks have initiated lively and partially controverse discussions. The special Dagstuhl atmosphere has helped a lot to develop a constructive atmosphere that improved mutual understanding and inspiration. The traditional hike on Wednesday afternoon and time spent together in the evenings or late at night let room for more personal discussions, too.

11.2 Perspectives Workshop: Quantum Computing

Seminar No. **04202** Date **09.05.–13.05.2004**
Organizers: T. Beth

In addition to the by now famous results of quantum algorithms and quantum cryptography, many new applications of quantum information processing suggest that a new area Quantum Informatics is emerging. This Dagstuhl Perspectives Workshop will bring together leading researchers whose contributions are likely to be essential to the new field. Another aim of this gathering will be to generate a vision for the future direction of research in this new area and to identify strategies to further the development of this field, internationally.

Large quantum computers, new quantum algorithms, and understanding the emerging quantum complexity classes are some of the guiding problems of Quantum Informatics. But Quantum Informatics also influences classical computer science as it shows e. g. limits of classical integration, leads to the development of new public key cryptography, and changes the interpretation of statistics. Quantum Informatics has immediate applications like quantum key exchange and applications which will be realized in the near future like quantum repeaters, entanglement purification, and quantum state stabilization. Furthermore Quantum Informatics with its aspect of quantum information has started to have a great impact on physics yielding a better understanding of quantum mechanics as well as applications within physics like improved measurement and control of quantum systems. On the other hand, quantum bounds may provide new insights about limits of computation in general.

The Dagstuhl Perspectives Workshop will hopefully help to propagate this promising field into the informatics community and thus constitute Quantum Informatics as a field sui generis at a broader scope, by identifying future perspectives, enlarge the number of researchers, and enhance collaboration between the different research groups working on quantum informatics.

The workshop is envisaged to concentrate on those aspects of Quantum Informatics which generalize the informatics concepts rather than dealing with experimental details of physical systems. To make this workshop a success, we have carefully selected a group of key researchers whose work lies in one or more of the following areas:

Quantum Computing:
novel algorithms, fault tolerant computing, quantum learning, impacts of quantum computing on classical cryptography and computing.

Quantum Communication:
novel protocols for quantum cryptography, security analysis of concrete prototypes, quantum repeaters, preparation of distributed entanglement.

Distributed Quantum Systems:
distributed quantum computation, applications of distributed entanglement, quantum secret sharing, secure multiparty quantum computations.

Quantum Information and Complexity Theory:
quantum limits of classical computations, quantum complexity theory, quantum communication complexity, quantum error-correcting codes, a complexity theory of quantum state preparation, quantum data mining, quantum statistical effects and causality.

Technological Applications:
quantum state preparation, quantum control, improved measurement techniques, entanglement enhanced lithography, enhanced microscopic imaging, quantum circuits, quantum memory, algorithmic cooling, non-standard quantum hardware.

11.3 Integrative Bioinformatics – Aspects of the Virtual Cell

Seminar No. **04281** Date **04.07.–09.07.2004**
Organizers: J. Collado-Vides, R. Hofestädt, C. Sensen

Molecular biology produces huge amounts of data in the post-genomic era. Among them, there is data describing metabolic mechanisms and pathways, structural genomic organization, patterns of regulatory regions; proteomics, transcriptomics, and metabolomics data. On one hand, analysis of these data is determined essentially by the methods and concepts of computer science; on the other hand, it depends on a range of biological tasks solved by researchers. Currently, there are about 500 database informational systems and various analytical tools available via the Internet and directed at solving various biological tasks. The challenge we have is to integrate these list-parts from genomics and proteomics at novel levels of understanding. Integrative Bioinformatics would be this new area of research using the tools of computer science and electronic infrastructure applied to Biotechnology. These tools will also represent the backbone of the concept of virtual cell.

Goals

Regarding the development of methods and concepts of Bioinformatics to model metabolic processes the integrative aspect stands in the center of the seminar. The discussion of specific requirements for the implementation of a virtual cell using these technics is the main goal of the seminar. Therefore, one result will be to characterize the possibilities of Integrative Bioinformatics within the scope of System Biology. This seminar addresses primarily scientists working in this emerging field of Integrative Bioinformatics and, hence, would discuss the corresponding problems and present concepts. The key idea is to invite scientists which are working in the dry and/or wet lab.